

The Infrastructure of Chimpanzee Vocal Ontogeny

Derry Taylor



Thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth.

December 2020

Abstract

Almost all living things communicate, yet only humans have language. The question of how this came to be has puzzled scientists and philosophers for centuries. In more recent years, researchers have looked to primate communication systems in order to gain insights into the evolutionary origins of language. While many studies have successfully identified language-like features in adult primate communication systems, the parallels between human and non-human primate vocal ontogeny are poorly understood. In this thesis, I aimed to address this issue by examining the process of vocal ontogeny in chimpanzees. In my first empirical study, I examined the ontogeny of the acoustic structure of the vocal repertoire. It was found that the chimpanzee vocal repertoire in the first 10 years of life did not increase in the number of call types, but became increasingly acoustically graded. In my second empirical study, I examined the ontogeny of patterns of vocal production and function, finding that from infancy chimpanzee grunts express a wide range of affective states, and later during the juvenile period these calls show signs of functional flexibility. Such patterns were not observed for any other call types. Finally, in my third empirical study, I examined the ontogeny of directedness and engagement during vocal communication. It was found that chimpanzees routinely showed directedness and engagement during vocal communication, directedness generally increased during ontogeny, and was associated with a higher probability of eliciting responses from social partners. Overall, the findings of this thesis suggest that while acoustically speaking, chimpanzee vocal ontogeny is rather different to human vocal ontogeny, chimpanzee vocal ontogeny is characterised by communicative capacities that are important precursors for language ontogeny (i.e. flexible vocal production, functional flexibility, directedness, and engagement). In turn, this might suggest such capacities were also important phylogenetic precursors to language.

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word count: 41,112

Acknowledgements

Firstly, I would like to thank my first supervisor, Marina Davila-Ross. From Marina I learnt one of the most important skills I will probably ever acquire in my career – how to see the fascinating things that animals do. Marina’s continued support and patience throughout the years of my PhD have allowed me to become a more competent and independent researcher. Most importantly for me, Marina’s calm demeanour, open-mindedness, and sense of humour, have made doing a PhD something very enjoyable to do.

I would also like to thank my second and third supervisors – Guillaume Dezecache and Erik Gustafsson. Guillaume’s own research was foundational in the development of this thesis and consequently his insights have been invaluable. I thank Guillaume for spending an enormous amount of time and effort sharing his expertise with me throughout the years of my PhD and for treating me as a colleague as well as a student. In doing so, Guillaume has provided me with an indispensable foundation to build my academic career upon. Erik, my third supervisor, has consistently and punctually provided feedback and support on my empirical projects. Erik’s knowledge of developmental psychology was also essential in this endeavour and enabled this project to fulfil its aim of finding parallels in human and chimpanzee vocal ontogeny. I am hugely grateful to Erik for this.

Beside my PhD supervisors, the hard work and support of many others have made this project possible. I would like to thank all of the staff at Chimfunshi Wildlife Orphanage for their support both as colleagues and friends during the field work. In particular, I would like to thank Tom, Rogers, Richard, Felix, Dennis, and Mishek. I would also like to thank my research assistants, Peter Kidd, Sarcha Wheeler, and Abbie Hitchenson who travelled from

the UK to Zambia to work on the project for several months. The fieldwork would not have been possible without their help, and their friendships.

Before my 20's, nobody who knew me (including myself) would have expected I would go on to earn a PhD. I had very little direction and interest in the world. Studying allowed me to see the beauty and complexity in the world around us. I thank my family for their unconditional love and support throughout the years which taught me that what matters is finding something that makes you happy. I finally feel I have found that thing.

Lastly, I would like to thank my partner, Jemma. Aside from putting up with me for almost a decade, she spent a year with me in Scotland while I earned my masters degree and then returned to Portsmouth for 3 years with me while I carried out my PhD work. During that time, she also spent 4 months living alone while I carried out field work. Jemma inspired me to study and helped me realise that there might be more out there for me. Without her support, I am certain that I would not be where I am today. I am immensely grateful for everything she has done for me throughout the years.

Dissemination of research from this thesis

Publications, abstracts, presentations, and posters

Chapter 2

- Taylor D., Dezechache, G., & Davila-Ross, M. (2020). Filling in the gaps: acoustic gradation increases during chimpanzee vocal ontogeny. *American Journal of Primatology*. Under review.
- Taylor, D., Dezechache, G., & Davila-Ross, M. (2020). Filling in the gaps: acoustic gradation increases during chimpanzee vocal ontogeny. Abstract at IPS international congress, Quito, Equador.

Chapter 3

- Taylor, D., Dezechache, G., Gustafsson, E., & Davila-Ross, M. (2018). The Development of Vocal Functional Flexibility in Semi-Wild Chimpanzees. Poster presentation at PSGB Spring meeting, Portsmouth, UK (April 11th – 12th).
- Taylor D., Dezechache, G., Gustafsson, E., & Davila-Ross, M. (2018). Do Young Chimpanzees Vocalise Freely?. Poster presentation at IPS international congress, Nairobi, Kenya (August 19th-25th).

- Taylor D., Dezecache, G., Gustafsson, E., & Davila-Ross, M. (2019). Free Vocal Expressivity in Young Chimpanzees. Oral presentation at EFP international conference, Oxford, UK (September 8th-11th).

Contents List

1. Introduction	1
Overview	1
Philosophy of language in the context of language origins	5
Insights into language evolution from comparative communication in primates.....	11
A comparative-developmental approach to language evolution.....	18
The infrastructural natural logic model	21
 2. Filling in the gaps: acoustic gradation increases in the vocal ontogeny of chimpanzees	 28
Abstract	28
Introduction	29
Method.....	33
Subjects and study site	33
Data collection	36
Identifying calls	37
Acoustic analysis	41
Statistical analysis	45
Results	51
Cluster validity, reliability, and stability.....	51
Cluster composition	54
Cluster typicality	55
Discussion.....	58
 3. Functional flexibility in chimpanzee vocal ontogeny.....	 66
Abstract	66
Introduction	67
Method.....	71
Subjects and study site	71
Data collection	74
Identifying calls	75
Coding and classifying subject behaviours of valence	76
Measuring social partner behaviour	80
Statistical analysis	82

Results	84
Discussion.....	88
4. Directedness and engagement in chimpanzee vocal ontogeny	97
Abstract.....	97
Introduction	98
Method.....	103
Subjects and study site	103
Data collection	105
Identifying calls	107
Directedness and engagement coding	108
Social partner behaviour and responses.....	110
Statistical analysis.....	111
Results	115
Ontogeny of directedness & engagement during vocal communication.....	115
Function of directedness & engagement during vocal communication.....	118
Discussion.....	119
5. General Discussion.....	127
Overview	127
Main findings and implications	130
Infraphonology of chimpanzee vocal ontogeny.....	130
Infrasemiotics of chimpanzee vocal ontogeny.....	139
Infrastructural natural logic model & the evolution of language.....	148
Overall conclusions and future directions	168
References.....	170
Appendices.....	189
Appendix 1 – Supplementary methods and results (Chapter 2)	189
Supplementary methods	189
Supplementary results.....	190
Appendix 2 – Supplementary results (Chapter 3)	213
A) Direct comparison of expression patterns between infants and juveniles.....	213
B) Direct comparisons of call functionality between infants and juveniles and functional flexibility analyses for laughs, screams, and whimpers	214
Appendix 3 – Supplementary methods & results (Chapter 4).....	222

Supplementary methods	222
Supplementary results.....	233
Appendix 4 – AWERB ethical approval.....	237
Appendix 5 – Chimfunshi Wildlife Orphanage research approval	238
Appendix 6 – UPR16 Form	240

List of tables

TABLE 2.1 DEMOGRAPHIC CHARACTERISTICS OF SAMPLE.....	31
TABLE 2.2 CALL TYPE DEFINITIONS.....	36
TABLE 2.3 MEASURED ACOUSTIC PARAMETERS AND DEFINITIONS.....	39
TABLE 2.4 CALL COMPOSITION OF EXTRACTED CLUSTERS.....	52
TABLE 3.1 DEMOGRAPHIC CHARACTERISTICS OF SAMPLE.....	69
TABLE 3.2 AFFECTIVE STATE CUES AND DEFINITIONS.....	74
TABLE 3.3 SOCIAL PARTNER BEHAVIOUR DEFINITIONS.....	78
TABLE 4.1 DEMOGRAPHIC CHARACTERISTICS OF SAMPLE.....	101
TABLE 4.2 SOCIAL PARTNER BEHAVIOURAL RESPONSES.....	107

List of figures

FIGURE 1.1 INFRASTRUCTURAL NATURAL LOGIC MODEL.....	20
FIGURE 2.1 SPECTRAL MEASUREMENTS.....	41
FIGURE 2.2 CLUSTER RELIABILITY FOR INFANTS AND JUVENILES.....	49
FIGURE 2.3 CLUSTER TYPICALITY FOR INFANTS AND JUVENILES.....	54
FIGURE 2.4 SPECTROGRAMS DISPLAYING TYPICAL CALLS FROM EACH CLUSTER.....	55
FIGURE 3.1 FREE VOCAL EXPRESSIVITY AND FUNCTIONAL FLEXIBILITY.....	82
FIGURE 4.1 MAIN EFFECTS AND INTERACTIONS FOR BEST FITTING MODELS OF DIRECTEDNESS AND ENGAGEMENT.....	113
FIGURE 4.2 RELATIONSHIP BETWEEN DIRECTEDNESS AND SOCIAL PARTNER RESPONSES.....	116
FIGURE 5.1 SUGGESTED REVISION OF THE INFRASTRUCTURAL NATURAL LOGIC MODEL.....	149

1. Introduction

Overview

Almost all living things communicate, yet only humans have language (Scott-Phillips, 2008). This simple observation has puzzled scientists and philosophers for centuries. This problem has even been viewed as so empirically intractable, that in 1866, 7 years after Darwin first published *On the Origins of Species*, the Linguistic Society of Paris banned all discussion on the evolutionary origins of language (Hauser et al., 2014). In the following century, great advances in phylogenetics revolutionised our understanding of the evolutionary relationships between living things (Martins, 1996). With an improved understanding of the evolutionary relationships between species, it became possible to reconstruct the origins of traits that do not fossilise such as language by comparing the behaviour of closely related species (Fitch, 2005). Consequently, the subject of the evolutionary origins of language was re-introduced into science.

Today, researchers typically attempt to gain insights into the evolutionary origins of language in a three-stage process (Fitch, 2000). First, key features of language such as syntax and semantics are identified. Second, researchers design studies to test whether these features are also characteristic of communication in nonhuman primates (from here on 'primates') closely related to humans. Finally, evolutionary inferences are made. If a given feature is shared by humans, and primates they are closely related to, we infer that this ability is rooted in our primate ancestry. If the capability in question is not found, we infer that this ability is likely uniquely human.

While a great deal has been learned about language origins by adopting this approach, some have said it is crucially limited in that it overlooks the importance of ontogeny (Rosati et al., 2014; Griebel et al., 2016; Oller et al., 2016). Since language emerges from a complex developmental process (Freeman & Cameron, 2008; Kretzschmar & Kretzschmar, 2015), this suggests that to understand the evolutionary origins of language we need to modify the mainstream comparative approach, by comparing not just overall features of communication systems, but how those systems develop (Griebel et al., 2016). While language can be implemented in the gestural domain, language is spoken by default (Sandler & Lillo-Martin, 2006). This is culturally universal – there is no known culture wherein the default language modality is gesture (Aaron & Joshi, 2006). As such, spoken language appears to be deeply rooted in human biology (Lieberman, 2006), and was the focus of this thesis. Given the importance of developmental processes in the vocal domain for our understanding of language origins, I aimed to adopt a comparative-developmental approach to language evolution, by examining the process of chimpanzee vocal ontogeny, and assessing whether this process bears any similarities with human vocal ontogeny.

In this first chapter, I will first discuss different philosophical conceptions of language and evaluate their influence on the way in which the question of language origins is approached and understood, concluding that language is built upon pragmatic foundations, but primatological research on vocal communication is largely ambivalent about the philosophical conceptualisations of language to which it is committed. I will then review and critique the mainstream comparative approach to language origins in primatological research and argue for the need to adopt a comparative-developmental

approach. For a comparative-developmental approach to language origins to progress, a model of vocal ontogeny that is suitable for cross-species comparisons is needed. I will go on to suggest that the infrastructural natural logic model of communicative development (Oller, 2012) represents such a model and finally review this model with respect to human infant and primate vocal communication literature.

The infrastructural natural logical model specifies ontogenetic changes in both vocal acoustics and vocal production that are essential as foundations for language. In the first empirical chapter, chapter 2, I provide some of the first systematic insights into the process of chimpanzee vocal ontogeny from an acoustic perspective by examining ontogenetic changes in the size and acoustic structure of the vocal repertoire within the first 10 years of life, which covers infancy through to sub-adulthood in chimpanzees (Goodall, 1990; Plooij, 1984). In human vocal ontogeny, a small number of highly graded call types is elaborated into a larger repertoire of more discrete sounds (Oller, 2000; Stark, 1981; Vihman, 2014). In chimpanzees, we find no increase in vocal repertoire size within the first 10 years, although we observed a significant increase in acoustic gradation between call types throughout ontogeny. This suggests that, acoustically speaking, chimpanzee vocal ontogeny is unlike human vocal ontogeny.

Chapter 3 focuses on ontogenetic changes in patterns of vocal production and function. Theoretically, the infrastructural natural logic model posits that speech and language emerge from vocalisations that are flexible both in terms of what they can express and the functions they can fulfil (Oller, 2012). This is supported by observations of human vocal ontogeny wherein the precursors to the first speech sounds in human vocal

ontogeny are freely expressed and functionally flexible, where the former refers to calls that express different affective states on different occasions, and the latter refers to calls whose function depends in part on how the call is expressed as opposed to what the call type is per se (Oller, 2012; Oller et al., 2013). Non-speech related vocalisations such as laughs and cries do not show such properties early in ontogeny (Oller et al., 2013; Jhang & Oller, 2017). I evaluated whether infant and juvenile chimpanzee vocalisations show signs of free vocal expressivity and functional flexibility. I found that grunts were not affectively biased, whereas laughs, screams, hoo calls, and whimpers were affectively biased. This was observed in both infants and juveniles. Grunts were also the only call type that met the criteria for functional flexibility – different types of responses were systematically elicited depending on the affective valence of the grunts. Juvenile grunts appeared to be more flexible with regard to function than infants. No other call types showed signs of functional flexibility.

According to the infrastructural natural logic model, freely expressed call types should also be more likely to be ‘freely directed’ (i.e. calls can be directed towards specific social partners by means of gaze or face directedness). Directed vocal behaviour and engagement during vocal communication (i.e. mutual gaze) are known to play an important role in human vocal ontogeny (Donnellan et al., 2020; Gros-Louis et al., 2014), but this has not been systematically studied in chimpanzee vocal ontogeny in a manner comparable to the human infant literature. In chapter 4, I therefore aimed to assess whether there are ontogenetic changes in directedness and engagement during chimpanzee vocal communication, whether this differs between call types, and whether this impacts social partner responses, as is observed in human vocal ontogeny. Similar to

human infants (e.g. Donnellan et al., 2020; Gros-Louis et al., 2014), I found that face directed vocalisations became increasingly common during ontogeny, and were also more likely to elicit behavioural responses from social partners, although the effect of age on gaze directed vocalisations and mutual gaze was dependent on call type and such calls were not more likely to elicit behavioural responses from social partners. The observations were also inconsistent with the predictions of the infrastructural natural logic model - screams and whimpers were generally most likely to be directed towards a social partner even though they are not freely expressed call types, and there was no significant difference between call types in how flexibly they were directed towards social partners.

In the final chapter, I summarise my findings both within the context of the infrastructural natural logic model and the study of language evolution more generally. I conclude that the infrastructural natural logic model identifies several pragmatic communicative capacities in early chimpanzee vocal communication that are known to be foundational in language development (i.e. free expressivity, functional flexibility, and the use of directed vocalisations to effectively engage with social partners), suggesting these characteristics may be foundational for language phylogenetically as well ontogenetically. However, given that many predictions of the infrastructural natural logic model were not supported, I argue this model may need slight revisions in order to accurately describe the *process* of chimpanzee vocal ontogeny.

Philosophy of language in the context of language origins

Throughout the majority of intellectual history, language has been studied predominantly by philosophers. Of particular interest amongst philosophers has been the attempt to develop a conceptual foundation upon which a full understanding of language can be built. In his 1921 book, *Tractatus-Logico-Philosophicus*, philosopher and logician Ludwig Wittgenstein argued for a 'picture theory' of language, whereby language is understood as a system of labels that are used to represent and communicate about a reality that exists prior to language. Wittgenstein later argued that the so-called 'picture theory of language' could not account for the variety of things that speakers do with language in social interactions, and in his subsequent book '*Philosophical Investigations*' (1953) argued instead for a pragmatic conceptual foundation for understanding language. In this later philosophical work, language is understood not as a set of labels with a one-to-one correspondence with objects in the world, but as a system that has meaning only in relation to speakers and listeners within a context.

In more recent history, the origins of language has become an object of scientific enquiry (Hauser et al., 2014). The way in which scientists approach the question is heavily influenced by underlying philosophical assumptions about the fundamental nature of language. For example, primatological research on so-called 'referential' communication (or the quest for word-like features in primate vocal communication) has clearly embraced a set of philosophical assumptions that align with Wittgenstein's picture theory of language. In their seminal paper on alarm calls among vervet monkeys, Seyfarth et al. (1980) indeed demonstrated that acoustically distinct vervet alarm calls were associated with different predator classes and elicited different behavioural responses from listeners that were suitable to avoiding the predator associated with the alarm call (but see Price

et al., 2015). The authors of this paper interpreted this pattern as evidence of 'functionally referential' vocal communication – calls refer to predators in a similar way to how words can refer to objects in the world. The interpretation of vervet alarm calls as referring to particular predator classes clearly aligns with Wittgenstein's picture theory of language whereby language is conceptualised as a system of labels with a one-to-one correspondence with objects in the world. This influence has since extended further into the primate communication literature, with subsequent studies showing evidence for so-called functionally referential vocal communication in a range of other species (macaques: Gifford et al., 2003; tamarins: Kirchhof & Hammerschmidt, 2006; chimpanzees: Slocombe & Zuberbühler, 2005).

While embracing a set of assumptions about the nature of language that aligns with the philosophy of early Wittgenstein has apparently been met with some success in primatology, some have pointed out crucial limitations. Critics have argued that seemingly functionally referential vocal behaviour can be explained without appeal to referentiality. For instance, signals may elicit behavioural responses from receivers without the signal informing the receiver about an object or event in the world (Owings & Morton, 1998; Owren et al., 2010). In fact, Oller & Griebel (2014) applied the notion of functional referentiality to crying among new born human infants. Citing evidence of acoustically differentiated cry-variants that elicit different behavioural responses from caregivers (e.g. Lester et al., 1995), Oller and Griebel (2014) argue that human infant cries within the first 6 months of life would meet the criteria for functionally referential vocal behaviour; in vervet monkeys acoustically distinct alarm calls elicit different behavioural responses from others much as acoustically distinct human infant cries elicit different

behavioural responses from others. Yet there are few who would maintain that human infant cries are referential in the sense that words and sentences are referential. Instead, Oller and Griebel (2014) argue that human infant cries do not *refer* to anything, but they *express* something (i.e. the state of the infant) which affects a listener in systematic ways.

The controversy surrounding functionally referential communication illustrates a point that originated with Wittgenstein but was largely developed in Austin's (1975) seminal book *How to do things with words*, in which Austin argued that communicating using language is an act – one does not simply say things that are either true or false, but instead when communicating with language one performs so-called 'speech acts'. Consider the sentence 'there's a car coming' uttered to someone about to cross a road. The sentence can be understood literally as a statement about the world. However, in addition to this, the sentence can be understood as an act, such as warning, which is intended to elicit a particular effect on a listener (i.e. that they move out of the way of the car). Thus, from an Austinian pragmatic perspective, what is most fundamental to linguistic communication is that what is communicated depends largely on what one *does* rather than literal information encoded into linguistic signals.

Those who study animal communication typically understand communication using the code model (Bradbury & Vehrencamp, 2011). In the code model, communication is conceptualised as a linear process whereby an organism produces a signal (say, a vocalisation), which travels through a medium (air, in the case of primate vocalisations), and another organism receives, decodes, and responds to that signal (Seyfarth et al., 2010). This approach to communication is rooted in Shannon's (1948) mathematical

theory of communication, wherein information is encoded in to signals based on statistical regularities in signal structure. Indeed, code-like systems of communication are common among animals (Scott-Phillips, 2015). However, as a development in communication systems engineering, Shannon's (1948) mathematical theory of communication was designed to model communication in closed systems (i.e. wherein signal meaning does not depend on anything external to the signal itself). As such, it has been argued that such systems are not comparable with human language (Oyama, 2000), wherein communication can only be understood with reference to the communicators behaviour, affective states, intentions, goals, and context, rather than meaning being encoded completely into the structure of signals (Wharton, 2003; Wilson & Sperber, 2012; Scott-Phillips, 2015).

Some have gone as far as to argue that understanding the origins of language is not possible from the code model perspective (Scott-Phillips, 2015). For example, in the code model, a signal has a prior (usually naturally selected) meaning (Maynard-Smith & Harper, 2003). However, Wittgenstein argued that if one assumes that meaning precedes communication, one cannot explain communication because one is unable to explain how shared meanings originate (Reddy, 2010). Similarly, Scott-Phillips (2015) argued that due to the indeterminacy of reference (i.e. linguistic utterances do not have intrinsic meaning in the same sense that naturally selected signals do), language is not possible without the ability make inferences about the communicative intentions of others. To understand the evolutionary origins of language, we must therefore go beyond the code model of communication and enquire into the origins of the pragmatic competencies that make language possible. Historically, primatological research has primarily adopted the code

model of communication and in many cases the code model appears to provide an adequate description of primate communication (Seyfarth et al., 2010). More recently, there has been growing support for the adoption of a pragmatic perspective in primate communication research based on observations of contextually flexible call production which is not befitting of the code model (Seyfarth & Cheney, 2018; Arnold & Bar-On, 2020). However, the question of whether primate vocal behaviour is also characterised by the kind of pragmatic qualities that define human vocal communication, which go far beyond contextually flexible vocal production, is less known.

In human ontogeny, pragmatic communicative competencies precede the linguistic capabilities that are more common targets of investigation in other species such as referentiality and syntax, and evidence suggests these competencies provide an essential foundation for these characteristics. For example, the function of the direct precursors to the first speech sounds depends on how those sounds are expressed (i.e. what behavioural indicators of affect they are accompanied by), which is not the case for non-speech related sounds. This implies communication depends largely on what infants *do* with their vocalisations rather than what the vocalisations are *per se* (Jhang & Oller, 2017). Additionally, infants within the first year who are better at engaging social partners through directed vocal behaviour, an important indicator of communicative intentions (Bruner, 1973), which are central for pragmatic communication (Scott-Phillips, 2015), show accelerated rates of language development (Donnellan et al., 2020). Since such pragmatic competencies appear to be foundational for language both theoretically and empirically in human vocal ontogeny, perhaps such pragmatic competencies provide the phylogenetic foundations for language to develop also.

Insights into language evolution from comparative communication in primates

To gain insights into the evolutionary origins of language primatologists look for evidence of language-like features in primate communication systems (Fitch, 2005). The first stage in this process is to identify key features of language, such as syntax and semantics.

Usually, the focus is on defining features of mature language. These features are then operationalised, and researchers design studies to test whether these features are also characteristic of communication in primate species. Finally, evolutionary inferences are made. If a given feature is shared by humans and primates they are closely related to, it is typically inferred that this ability is rooted in our primate ancestry. If it is not found, it is typically inferred that this ability is likely uniquely human and its emergence may have been a key evolutionary transition towards language (Fitch, 2000).

A central debate in the comparative literature on language origins is whether language is phylogenetically rooted in manual gesture or vocal forms of communication (Corballis, 2017). By default, human language is vocal. While some primate calls (e.g. great ape laughter: Davila-Ross et al., 2009) are phylogenetically related to human non-verbal vocalisations, the phylogenetic relationship between speech sounds and primate vocalisations is poorly understood. It is commonly believed that due to the flexibility with which gestures are produced and various successful attempts to teach novel forms of gestural communication to great apes, that language is phylogenetically rooted in gesture rather than vocal communication (see Tomasello & Call, 2007). This view has been progressively undermined by more recent studies showing a greater degree of flexibility

in vocal communication in great apes than previously thought, both in terms of production and call acoustics (see Fedurek & Slocombe, 2011). A reconciliatory view is that language is rooted in multi-modal communication – language does not exist in isolation from other forms of communication (i.e. facial and gestural), which are often central to understanding vocal linguistic communication, and hence a multi-modal approach is essential for understanding language origins (Slocombe et al., 2011). In the present project, I make no commitment with regard to the modality within which language first emerged. I do contend, however, that comparative studies of vocal communication do provide insights into the evolutionary roots of language by providing information about what qualities vocal communication systems likely had when humans transitioned into vocal language, and consequently pointing out what qualities would have needed to emerge to facilitate that transition. In what follows, the focus will therefore be almost exclusively on vocal communication of great apes due to their phylogenetic closeness to humans (Prado-Martinez et al., 2013). In particular, I will explore the topics of referentiality, syntax, vocal learning, and pragmatics, which represent the key strands of research in this field to date.

Referentiality

The comparative approach has revealed much about the evolutionary origins of language. As mentioned above, many primate species have been shown to exhibit ‘functionally referential’ vocal communication, although this interpretation is controversial (see Fischer & Price, 2017). Aside from the previously discussed philosophical scepticism, many studies of apparently referential communication have focused on contexts wherein the referent is present. Humans, by contrast, often refer to absent referents (Hockett,

1960). A small number of chimpanzee gesture studies have claimed to find evidence for reference to absent objects (Lynn et al., 2014; Bohn et al., 2015), although this too has been subject to alternative interpretations (see Liszkowski et al., 2009). However, more recent studies of wild orangutans have shown that alarm calls were actually suppressed for up to 20 minutes after a predator left before being produced (Lameira & Call, 2018), providing the first evidence of reference to absent entities in natural great ape vocal communication.

Syntax

Studies on so-called referential communication in primates clearly suggest the capacity for such communication may have deep evolutionary roots in the primate lineage, and perhaps beyond (Townsend & Manser, 2013). However, research into the evolutionary roots on syntax do not provide such a clear picture. Experimental studies of many primate species, including Titi (Schlenker et al., 2017), Putty-nosed (Arnold & Zuberbühler, 2012), Diana (Zuberbühler, 2002), and Campbells (Ouattara et al., 2009) monkeys show that these species produce signal combinations that have functions that are distinct from the functions of the constituent calls in the sequences. Great apes too, are known to produce combinations of calls in a sequence. Chimpanzees, for instance, are widely known for producing pant-hoot series which are composed of four distinct phases (Slocombe & Zuberbühler, 2010), each of which encodes different information related to caller identity, social status, and context of production (Fedurek et al., 2016). However, there is limited evidence on the extent to which each phase is used outside of the production of pant-hoots, meaning it is difficult to ascertain the extent to which each unit is meaningful in isolation, in turn obscuring whether the whole is in anyway different to

the sum of its parts as it is in the syntactic call combinations of the aforementioned monkey species (Leroux & Townsend, 2020). Consequently, the extent to which human syntax is rooted in our shared ancestry with great apes remains unclear.

Vocal learning

Vocal learning is also considered to be a defining characteristic of language, leading primatologists to ask whether this is also a characteristic of primate vocal communication. Indeed, it has even been argued that in order to consider great ape vocal behaviour as homologous with human vocal behaviour and thereby relevant to the study of language evolution, vocal learning in great ape vocal behaviour must be demonstrated (Lameira, 2018). More specifically, it has been argued that great ape vocalisations that share articulatory characteristics with human speech sounds (i.e. vowels and consonants), should show patterns consistent with learning in order to be considered homologous with human speech sounds (Lameira, 2018).

Great apes show a range of unvoiced consonant-like sounds (Orangutans: Wich et al., 2009; Chimpanzees: Watts, 2015; Gorillas: Robins et al., 2016) and voiced vowel-like sounds (Orangutans: Lameira et al., 2016), which both experimental and observational research suggests can be modified through learning. Regarding consonant-like sounds, it has been shown that Orangutans are able to socially learn to produce unvoiced whistles from both conspecifics (Lameira et al., 2013) and human demonstrators (Wich et al., 2009), whose whistles they can also flexibly imitate in 'do as I do' imitation paradigms. Using this paradigm, it has also been shown that Orangutans can flexibly adapt the acoustic characteristics of novel voiced calls (Lameira et al., 2016), further suggesting that

they are socially learned. Other great apes such as chimpanzees have been observed to socially learn novel voiceless calls used for attention-getting by conspecifics (Tagliabata et al., 2012), which has in turn been shown to be associated with increased grey matter volume in Broca's area homologues (Bianchi et al., 2016) indicating a direct relevance to understanding language evolution due to the role of these areas in speech and language production, although further research on social learning of voiced calls in these species remain to be conducted. Nonetheless, these studies suggest that the vocal behaviour of great apes includes the basic building blocks for language – learned voiced and unvoiced calls.

While many of the aforementioned studies of great ape vocal learning occurred in captivity, observations of wild populations are also consistent with these studies. Two key forms of cross-population variation in learned vocal behaviour that are particularly interesting within the context of language evolution are dialects and accents. Dialects refer to population-specific words that share the same referent, for example, Americans use the word 'pants' for what English people would call 'trousers'. Similarly, observational studies of wild Orangutan vocal behaviour suggest the presence of regional dialects – different populations produce categorically different call types in the same context, similar to human synonyms (van Schaik et al., 2003; Krützen et al., 2011; Wich et al., 2012). For example, in some populations, individuals produce 'raspberries' during nest construction, while in others they produce 'nest smacks' (Wich et al., 2012). Importantly, these findings could not be explained by either genetics or ecology, suggesting such regional dialects are learned. While such synonym dialects have not yet been demonstrated among other great ape species, the binary presence vs absence of call

types has been demonstrated between individuals and populations of bonobos (Hopkins & Savage-Rumbaugh, 1991) chimpanzees (Watts, 2016), and gorillas (Robbins et al., 2016), indicating the possibility of dialects in other great ape species. Overall, these studies suggest that human dialects are homologous with those of great apes.

Similarly, in several great ape species there is evidence of accents – socially learned regional differences in the acoustic characteristics of the same call type. Marshall et al. (1999) studied the acoustic features of adult male pant hoots in two different captive populations in the USA, finding acoustic differences in call structure between these populations. Interestingly, the habitat and sub-species composition of these populations was highly similar, meaning differences in call acoustics could not be explained by differences in ecology or genetics, suggesting learning may be responsible. Moreover, in one population at Lion County Safari, Marshall et al. (1999) observed one male introduce an acoustically novel pant-hoot variant, which subsequently spread to five other males within the population, a pattern which the authors suggested was best explained by social learning. This interpretation is also consistent with earlier studies that showed a positive relationship between acoustic similarity of calls between male chimpanzees and the amount of time spent together (Mitani & Brandt, 1994). More recently, Watson et al. (2015) took advantage of a unique circumstance wherein two captive chimpanzee populations at were integrated at Edinburgh zoo. In an analysis of the acoustic structure of food-grunts before and after the integration, the authors found that the acoustic structure of the immigrant groups' food grunts became increasingly similar to those produced by the resident group, but only after social bonds had been established.

Together, these studies indicate that great ape vocal behaviour appears to show acoustic characteristics in common with the building blocks of human speech and language – voiceless and unvoiced call types that are capable of being acquired and modified through social learning. Such capabilities appear to give rise to natural linguistic-like phenomena such as accents and dialects across populations. Overall, this indicates the relevance of great ape vocal behaviour to understanding the evolutionary origins of language, both with regard to call production and call acoustics.

Pragmatics

Although primatology has been largely ambivalent about the philosophical concept of language that it embraces, studies of great ape vocal communication do indeed show evidence of pragmatic competency. From the Gricean perspective, pragmatic communication involves accounting for others' mental states during communication (Grice, 1957). Indeed, there is a growing body of research showing that great apes understand others' mental states including their knowledge states (Kaminski et al., 2008), and possibly even beliefs (Krupenye et al., 2016). Consistent with this, several field studies of chimpanzee vocal communication have shown that individuals selectively produce or suppress calls depending on the presence or absence of specific individuals (Schel et al., 2013; Slocombe & Zuberbuhler, 2007; Townsend & Zuberbuhler, 2009), suggesting chimpanzees consider whose mental states might be altered through their communicative acts. Further, several field studies have also shown that chimpanzees account for what others know in their alarm call production, producing more alarm calls when conspecifics were ignorant about the presence of a predator (Crockford et al., 2012, 2017). Thus, chimpanzees appear to account for others' mental states during vocal

communication, providing evidence that the capacity for pragmatic communication preceded the emergence of language. It is important to note, however, that this cognitive form of pragmatics is not identical to Austinian pragmatics wherein the focus is on the *action* of the communicator, rather than the communicators' understanding of other minds. Indeed, it is the Austinian variety of pragmatic competencies that appear to provide a pre-verbal foundation for the emergence of language (Griebel et al., 2016) and have been far less studied in primate communication.

Summary

Overall, there is evidence of a range of language-like features in the vocal communication systems of great apes and other primates, including referentiality, syntax, vocal learning, and pragmatic usage of communicative signals. This suggests that the capacity for language may be to some extent rooted in the more primitive communication systems of our primate ancestors. Although, it is important to note, that there is a clear ambivalence in primatology about the philosophical conception of language to which researchers are committed. This is important to address because the extent to which language-like communicative features appear to be present in the communication systems of other primate species largely depends on what type of a communication system we believe language to be.

A comparative-developmental approach to language evolution

Clearly, much has been learnt about the evolutionary origins of language through the mainstream comparative approach to language evolution. However, there are also many limitations of this approach aside from the aforementioned lack of a clear commitment to

a particular philosophical conceptualisation of what language is. One problem is that the focus on features of mature language overlooks many important precursors to these features. Defining characteristics of mature language (i.e. symbolism, phonology, syntax, etc.) do not appear *ex nihilo*, but are made possible by other prior capacities. For example, symbolic reference is made possible by the simpler yet logically prior capacity for indexical reference (Deacon, 1998), meaning is argued to be made possible by theory of mind (Scott-Phillips, 2015), and phonology requires fine-grained motor control of the vocal apparatus (Goldstein et al., 2006). The comparative approach to language evolution could therefore benefit by asking whether primate communication systems exhibit pre-linguistic non-verbal capacities that are logically necessary for the emergence of language (Oller, 2012).

Another, albeit related, issue is that the mainstream comparative approach has focused almost exclusively on adult primates (Bard & Leavens, 2014). However, language emerges from a complex developmental process involving pre-linguistic non-verbal capacities that are logically prior to language (Oller, 2000; Locke & Bogin, 2006). In evolutionary biology, it has now become widely recognised that natural selection sorts between not only different phenotypic endpoints, but also between developmental processes that give rise to those endpoints (Müller, 2007). The modification of developmental processes also provides a mechanism by which natural selection can add complexity to pre-existing structures and generate novel phenotypes (Brigandt & Love, 2010). Consequently, to understand the evolutionary origins of language an alternative approach is needed. In particular, the comparative approach to language evolution could benefit from comparing not just overall features of communication systems, but how those systems develop in

ontogeny (Griebel et al., 2016). However, primate vocal ontogeny is heavily understudied, even in more extensively studied and phylogenetically relevant species such as chimpanzees.

Some of the earliest descriptions of the vocal repertoire of infant chimpanzees were reported by Frans Plooij in 1984. In a study of wild infant chimpanzees in Gombe National Park, Plooij (1984) classified 22 types of vocalizations in chimpanzees up to 1 year of age. Plooij's (1984) call classifications mostly consisted of different types of grunts and whimpers, which is consistent with Kojima's (2008) analysis of infant chimpanzee vocalisations within the first 4 months of life using visual inspection of spectrograms that revealed two broad call categories – grunts and cries. Interestingly, Plooij (1984) did not report any evidence of key features of the adult vocal repertoire such as pant hoot calls, suggesting these emerge later in ontogeny. In a study of emotional development in neonatal chimpanzees, Bard (1998) reported observing threat barks and alarm calls within the first 19 days, and laughs within the first 37 days of life. More recent quantitative acoustic studies of infant chimpanzee whimpers found that whimpers could be reliably discriminated based on identity (Levréro & Mathevon, 2013) and revealed acoustically distinct whimper variants which were in turn associated with different maternal responses (Dezecache et al., 2020). Further, Dezecache et al. (2020) found infant grunts could be produced across a range of affective contexts and observed systematic acoustic variation in grunts according to context. While these studies provided some of the first insights into early chimpanzee vocal behaviour, they leave the ontogenetic gap between infant and adult vocal behaviour unexplored.

To our knowledge, only 2 studies have systematically explored chimpanzee vocal ontogeny beyond the infant period. Laporte & Zuberbühler (2011) studied the ontogeny of grunt production in infants, juveniles, and sub-adults, finding that chimpanzees decreased in the frequency of grunt production throughout ontogeny, but become more specific in who they direct their grunts to as they become older. In particular, grunts become more exclusively directed towards dominant males. More recently, Dezechache et al. (2019) studied alarm call ontogeny, finding appropriate use of alarm calls (i.e. production in the presence of a predator) increased with age. These studies show that chimpanzee vocal behaviour, at least with regard to vocal production, does change during ontogeny. However, there is a clear lack of a coherent research programme that aims to elucidate whether the process of chimpanzee vocal ontogeny shows signs of the pre-linguistic non-verbal communicative capacities that provide the necessary infrastructure for the emergence of language. As such, an important step towards advancing our understanding of the evolutionary origins of language is adopting a comparative-developmental approach focusing on similarities and differences in chimpanzee and human vocal ontogeny.

The infrastructural natural logic model

A comparative-developmental approach to language evolution is clearly needed. However, to achieve this goal, we first need a coherent model that identifies the relevant features of developing communication systems that provide the necessary foundations for the emergence of language. Oller (2012) developed a framework that fulfils this need. Oller's (2012) 'infrastructural natural logic' (IFNL) model is built on Ferdinand de Saussure's (Saussure, 1989) argument that to understand language one must distinguish

between the ‘signifier’ (i.e. the signal) and the ‘signified’ (i.e. the content). As such, the IFNL model specifies a set of ‘infraphonological’ features which pertain to signal structure and complexity, and ‘infrasemiotic’ capacities which pertain to signal meaning and function. Together, these capacities are argued to describe the path from early vocal behaviour to language development (Figure 1.1).

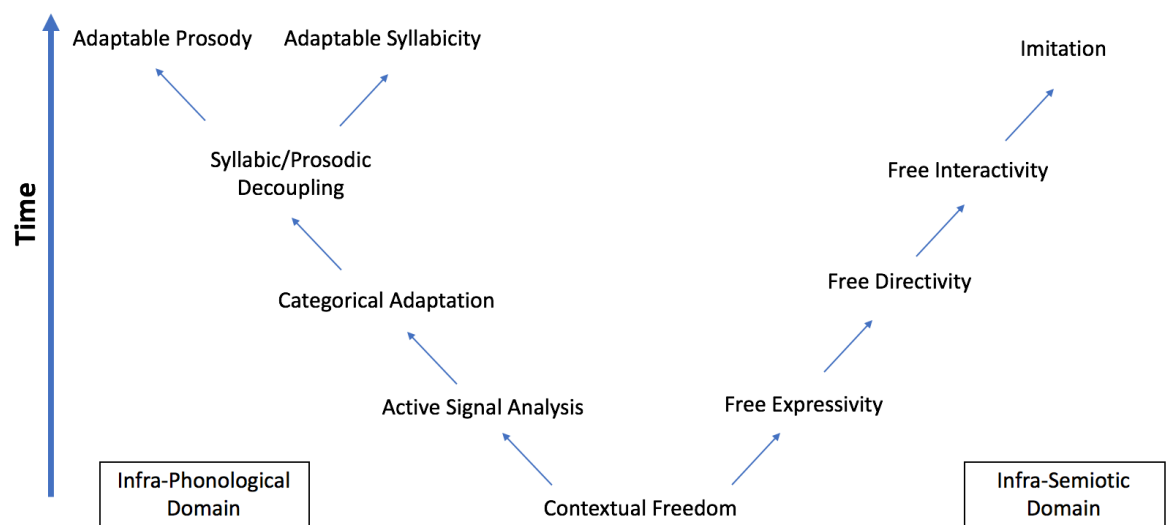


Figure 1.1 The infrastructural natural logic model of vocal communication (reproduced from Oller et al., 2012)

The capacities that are argued to be essential for language in the IFNL model are inspired by the work of Hockett (1960), who outlined a set of 16 features that define linguistic communication systems. Hockett’s (1960) key insight was that in order to compare language with other communication systems, the units of comparison cannot be concrete units of language, but instead must be ‘design features’ that make concrete units of language possible. For example, the unit of comparison cannot be a particular word such as ‘cat’, but instead must be a design feature that makes the emergence of such words

possible at all (i.e. the capacity for symbolic reference, socially learned communication, the ability to produce phoneme-like units and combine them into higher-order units, etc). Despite its valuable insights, Oller (2012) argued Hockett's (1960) framework overlooked the relationship between design features, and reviews on the empirical evidence suggested that Hockett's (1960) model could not clearly discriminate between human language and primate communication systems (Snowdon, 2012). Oller's (2012) IFNL model thus embraces Hockett's (1990) insight that relevant units of comparison are design features. However, the IFNL model builds upon Hockett's work by outlining the order in which these capacities emerge (Figure 1.1).

According to Oller (2012), these capacities emerge in the order that they emerge in because the emergence of a capacity at any given stage is made possible by the pre-existing capacities and therefore logically follows from them. As such, the IFNL model makes explicit the relationship between design features. Importantly, the IFNL model also identifies non-verbal communication as relevant to language because language according to the IFNL model is built upon pre-existing (and therefore non-verbal) communicative infrastructure. Since human non-verbal vocalisations have phylogenetic precursors in the great ape lineage (Davila-Ross et al., 2009; Schwartz et al., 2020), the IFNL therefore identifies great ape vocal communication as relevant targets for a comparative-developmental enquiry into the evolutionary origins of language. This is further supported by a recent comparative study of infant bonobo's which reported bonobo infant vocalisations have acoustic properties that are similar to human infant pre-speech sounds known as protophones (see Oller et al., 2019).

In primitive communication systems, according to Oller (2012), signal structure and function are tightly coupled. For example, acoustically distinct vervet alarm calls reliably elicit different responses from conspecifics (Seyfarth et al., 1980), suggesting the signal structure is tightly coupled to signal function. By contrast, language is marked by a decoupling between signal structure and function – the relationship between the signifier and the signified is arbitrary. For example, the word ‘burro’ means ‘donkey’ in Spanish, but means ‘butter’ in Italian, demonstrating a decoupling of signal and function. Thus, according to the IFNL model, signal-function decoupling, which is exemplified in its simplest form by context-free vocal production, is a fundamental design feature without which languages could not emerge (Oller, 2012). Contextually free calls, according to the IFNL model, will be more flexible regarding the acoustic form of the calls, due to contextual variation in factors such as arousal and affective state, and will also be more flexible with regard to what signallers are able to do with those calls, because signal meaning is not limited by context (Oller, 2012). It is important to note that this is clearly consistent with the pragmatic notion that communication depends on what communicators do with their communicative acts rather than on information contained within signals alone (Austin, 1970) because the behaviour of a communicator can provide information about what they are communicating when context or signal alone is insufficient due to variation in contextual usage or signal characteristics.

Indeed, such pragmatic competencies appear to be fundamental to the emergence of language in human vocal ontogeny. Protophones, which are believed to be the non-verbal precursors to the first speech sounds and therefore fundamental to language development (Yoo et al., 2018), are produced in a context free manner within the first

month of life (Jhang & Oller, 2017). From the 3rd month, protophones also show flexibility in what they are used to express and their effects on social partners depend on how they are expressed (Oller et al., 2013). This is not found among other human infant calls such as cries and laughter, which are more stereotyped in what they express (i.e. negative and positive affective states respectively) and their effects on social partners (Oller et al., 2013). This latter capacity to flexibly express different affective states with the same call type is known as 'free expressivity' in the IFNL model and the ability to harness this flexibility to achieve different functions with the same call is known as 'functional flexibility' (Oller, 2012). These capacities are argued to logically depend on the capacity for contextually free vocal production, because if a call is contextually limited this also limits what a signaller can potentially do with that call. For example, a contextually limited call like a vervet snake alarm call cannot be used to communicate about food location. Once human infant protophones are freely expressed they can also be freely directed (i.e. a socio-communicative target is selected using gaze and facial orientation) which is predicted by the IFNL model (Oller, 2012), and is known to be a pragmatic competency that plays an important role in language development (e.g. Donellan et al., 2020). Acoustically, protophones are more variable in form than other early vocal types, and are elaborated into a wider set of discrete calls which provide the 'raw material' for language in early human ontogeny (Stark, 1981; Vihman, 2014).

Since the IFNL model appears to accurately describe early ontogenetic changes in human non-verbal vocal communication that provide a foundation for the emergence of language, it may also provide a useful framework for a comparative-developmental enquiry into the evolutionary origins of language. In the primate literature, many species

are known to show contextually flexible vocal production; the grunts of adult baboons (Maciej et al., 2013; Rendall et al., 1999), coo calls of adult Japanese macaques (Koda, 2004), grunts in adult and juvenile mangabeys (Range & Fischer, 2004), adult bonobo high hoots (Schamberg et al., 2016) and peeps (Clay, Archbold, & Zuberbühler, 2015), and of particular relevance in the present case, adult chimpanzee grunts (Slocombe & Zuberbühler, 2010). Phylogenetically, such calls could provide a non-verbal foundation for the emergence of language, according to the IFNL model which posits context-free vocal production is an essential design feature for the emergence of language. However, no studies of primate communication have tested the predictions of the IFNL model. As such, in this thesis, I aim to assess the extent to which the IFNL model accurately describes the process of chimpanzee vocal ontogeny.

On the infraphonological side, in the first empirical chapter (Chapter 2), I examine ontogenetic changes in the size and structure of the chimpanzee vocal repertoire in order to evaluate whether a larger repertoire of more acoustically discrete call types emerges in ontogeny. This is known as ‘categorical adaptation’ in the IFNL model (Figure 1.1) and is routinely observed in human vocal ontogeny, but is more commonly known as the ‘expansion stage’ (Stark, 1981; Vihman, 2014). On the infrasemiotic side, in the second empirical chapter (Chapter 3) I test for the presence of free vocal expressivity and functional flexibility, as is seen in early human vocal behaviour (see Oller et al., 2013). Finally, in the third empirical chapter (Chapter 4), I examine the ontogeny of ‘directed’ vocal behaviour (i.e. the use of gaze and face orientation to select specific social targets) and engagement during vocal communication. Vocal directedness and engagement are known to play an important role in human vocal ontogeny (Donnellan et al., 2020), and

according to the IFNL model, these qualities made possible only by the capacity for free expressivity (Figure 1.1). Together, these studies will provide insights into the phylogenetic roots of the early non-verbal communicative infrastructure that provides an essential foundation for the emergence of language.

2. Filling in the gaps: acoustic gradation increases in the vocal ontogeny of chimpanzees

Abstract

Human vocal ontogeny is considered to be a process whereby a large repertoire of discrete sounds seemingly emerges from a smaller number of more acoustically graded vocalisations. While adult chimpanzee vocal behaviour is highly graded, its developmental trajectory is poorly understood. In the present study, we therefore examined the size and structure of the chimpanzee vocal repertoire at different stages of ontogeny. Audio recordings were collected on infant (N=13) and juvenile (N=13) semi-wild chimpanzees at Chimfunshi Wildlife Orphanage, Zambia, using focal and ad libitum sampling. All observed call types were acoustically measured. These were predominantly grunts, whimpers, laughs, screams, hoos, and barks. A range of spectral and temporal acoustic parameters were extracted, and fuzzy c-means clustering was used to quantify the size and structure of the repertoire. The infant and juvenile vocal repertoires were both best described with the same number of clusters. However, compared to infants, juvenile call clusters were less distinct from one another and could be extracted only when a low level of overlap between call clusters was permitted. Moreover, the acoustic overlap between call clusters was significantly greater for juveniles. Overall, this pattern shows greater acoustic overlap in juvenile vocalisations compared to infants, suggesting a trend towards increased acoustic gradation in chimpanzee vocal ontogeny. This may imply in contrast to humans, chimpanzees become increasingly proficient in using graded signals effectively rather than developing a larger repertoire of more discrete sounds in ontogeny.

Introduction

To understand the evolutionary origins of language, researchers typically search for language-like features in adult primate communication systems (Fitch, 2005). While a great deal has been learned about language origins by adopting this approach, it is crucially limited in that it overlooks the importance of ontogeny. On the ontogenetic path towards language, a key acoustic change in vocal behaviour within the first year of life is the emergence of an increasing number of discrete vocal units from a smaller number of more acoustically graded vocal units. This transition, widely known as the expansion stage (Oller, 2012; Vihman, 2014), is a routine feature of typical human vocal ontogeny (Stark, 1981) and is accompanied by changes in vocal tract morphology that afford infants a wider acoustic space of possible sounds that they can produce (Lieberman, 1984; Crelin, 1987). Early vocal repertoire expansion has been theorised to function as a fitness indicator used to capture the attention of caregivers (Zuberbühler, 2012; Locke, 2017). This is argued to be beneficial in the human ontogenetic niche due to high levels of co-operative breeding and alloparental care (Zuberbühler, 2012; Locke, 2017). Support for this argument comes from studies that show caregivers consider infants with more elaborate vocal behaviour to be more likeable (Bloom, D'Odorico & Beaumont, 1993; Bloom & Lo, 1990) and are more responsive to infants with a more elaborate vocal repertoire (Goldstein & West, 1999).

There is evidence of call types entering and leaving the vocal repertoire during primate vocal ontogeny (Gautier, 1974; Lieblich et al., 1980). However, whether an expansion in call types occurs at the level of the repertoire is less clear. In cotton-top tamarins, infants and juveniles produce only a single type of long call, whereas adults produce 3 long-call

variants (Snowdon, Cleveland & French, 1983), suggesting vocal-type expansion may occur during the ontogeny of tamarin long-calls. Among pigtailed macaques, misclassifications of scream call context based on call acoustics are significantly higher among infants than in older individuals (Gouzoules & Gouzoules, 1989), suggesting the development of a larger number of context-specific screams. This latter finding is consistent with vocal type expansion because older individuals may produce a wider range of scream types than infants. However, it could alternatively be that call types simply become more clearly acoustically differentiated during ontogeny but call types *per se* do not expand (i.e. increase in number).

To understand the evolutionary origins of language, it is important to focus on species closely related to humans, such as chimpanzees (Prado-Martinez et al. 2013). Vocal type expansion is particularly difficult to assess in chimpanzees because the calls of adult chimpanzees are notoriously variable in their acoustic structures, or graded (Slocombe & Zuberbühler, 2010). Despite the gradedness of chimpanzee vocalisations, human listeners have been able to reliably identify particular calls of chimpanzees and link them to distinctive functions (e.g. Goodall, 1990). Moreover, chimpanzee vocal tract morphology shows ontogenetic changes such as descent of the larynx and hyoid bone deeper in to the throat that can in principle increase the range of possible sounds that can be produced (Nishimura et al., 2003, 2006). However, it is not known whether repertoire expansion occurs or whether the degree of acoustic gradation of vocalisations changes during chimpanzee vocal ontogeny, as is observed in human vocal ontogeny (Stark, 1981; Oller, 2012).

In fact, chimpanzee vocal ontogeny is a poorly understood process. Some of the earliest descriptions of the vocal repertoire of infant chimpanzees were reported by Frans Plooij in 1984. In a study of wild infant chimpanzees in Gombe National Park, Plooij (1984) classified 22 types of vocalizations in chimpanzees up to 1 year of age based on the researcher's perception (not based on acoustic analysis). Plooij's (1984) call classifications mostly consisted of different types of grunts and whimpers, which is consistent with Kojima's (2008) analysis of infant chimpanzee vocalisations within the first 4 months of life using visual inspection of spectrograms that revealed two broad call categories – grunts and cries. Interestingly, Plooij (1984) did not report any evidence of key features of the adult vocal repertoire such as pant hoot calls, suggesting these emerge later in ontogeny. However, to our knowledge there are no systematic acoustic studies of vocal behaviour focused on stages between infancy and adulthood (i.e. juveniles). Furthermore, all classifications that focus on the vocal repertoire of older individuals (i.e. sub-adults and adults) vary in estimates of the number of call types between 12 and 32 (van Hoof, 1973; Goodall, 1990; Clark, 1991). Such varying estimates of the chimpanzee vocal repertoire size create an unclear picture of whether vocal type expansion occurs during chimpanzee vocal ontogeny and may reflect methodological difficulties in identifying discrete call categories from acoustically graded vocal behaviour (Crockford, 2019).

Historically, acoustic gradation in vocal behaviour has been largely viewed as a methodological challenge to contend with due to the associated difficulties with identifying reliable vocal units that researchers agree upon (e.g. Crockford & Boesch, 2005; Slocombe & Zuberbühler, 2010). Instead, acoustic gradation can be viewed as a potentially functional feature of a communicative system that increases the information-

encoding potential of a signal (Fischer, Wadewitz, & Hammerschmidt, 2017) and may thereby open up new functional possibilities that could be beneficial in more complex social niches. This suggests that by measuring acoustic gradation rather than treating it as noise, researchers can better understand communication (Wadewitz et al. 2015, Fischer et al. 2017). To date though, very few researchers have adopted such an approach.

To our knowledge, such an approach has been adopted in primate communication research in just one study, which focused on predominantly adult chacma baboons whose vocal repertoire size has been contested in the literature due to high levels of acoustic gradation (see Wadewitz et al. 2015; Fischer et al. 2017). Wadewitz et al. (2015) applied both hard and soft (or fuzzy) clustering techniques to acoustic data extracted from baboon vocalisations. Hard clustering techniques identify a finite number of discrete clusters (MacQueen, 1967), whereas fuzzy clustering acknowledges that cluster membership is often imperfect and therefore allows a given observation to overlap partially with different clusters (Zadeh, 1965). This cluster overlap can be used to measure gradation. Using hard-clustering techniques, Wadewitz et al. (2015) found that up to 20 call types could be extracted but the algorithm did not show strong support for any of the sets of clusters that could be extracted, suggesting considerable acoustic gradation in the data. Consistent with this, when adopting a fuzzy-clustering approach, Wadewitz et al. (2015) found that the chacma baboon vocal repertoire was best described as comprising a small number of clusters (2-5) with considerable acoustic gradation between clusters.

In the present study, we aimed to examine ontogenetic changes in chimpanzee vocal repertoire size and acoustic gradation in order to evaluate whether vocal type expansion occurs during chimpanzee vocal ontogeny. We used a sample of 26 young semi-wild chimpanzees, ranging from neonatal infants to juveniles on the brink of sub-adulthood. Leaning on the protocol adopted to study the vocal repertoire of chacma baboons (Wadewitz et al. 2015; Fischer et al. 2017), we assessed the number of clusters that could be reliably extracted, the validity of those clusters, and derived measurements of acoustic gradation (hereafter ‘typicality’) for both infants and juveniles. If vocal type expansion occurs during chimpanzee vocal ontogeny as it does in human vocal ontogeny (Stark, 1981), we would expect to observe a higher number of reliable clusters in juveniles, those clusters would be expected to have higher cluster validity, and they would be expected to be significantly less graded.

Method

Subjects and study site

Subjects were infant (N=13) and juvenile (N=13) semi-wild chimpanzees housed at Chimfunshi Wildlife Orphanage, Zambia. Infant ages ranged from 0 up to 4 years of age ($M=1.13 \pm SD=1.14$). Individuals aged between 4 – 10 years of age were classified as juveniles ($M=7.07 \pm SD=3.82$). These age ranges are the broadest definitions of these developmental stages in the chimpanzee literature and were therefore chosen to give the most comprehensive insight into chimpanzee early vocal development. The infants comprised 6 females and 7 males, whereas juveniles comprised 4 females and 9 males. Typical infant characteristics include riding either ventrally or dorsally with the mother and breast-feeding from the mother. Typical juvenile characteristics include no longer

riding with the mother either ventrally or dorsally, less reliance on breast feeding from the mother, and more independence from the mother (See Reynolds, 2005; van de Rijt-Plooij, & Plooij, 1987; Laporte, & Zuberbühler, 2011). An individuals' status as infant or juvenile in this study was, however, defined based on age alone.

Subjects are known to belong to a mixture of sub-species including *Pan troglodytes troglodytes* and *Pan troglodytes schweinfurthii*. However, many sub-species memberships were not known, meaning it is possible that the sample contained members of other sub-species. All subjects were raised by their mothers during infancy. All infant subjects still lived with their biological mothers. Three juveniles did not live with their mothers due to fatalities that occurred in years prior to the present study. We do not believe this impacted the data in the present study because chimpanzee orphans have only been shown to differ from typically reared individuals in terms of social behaviour and cognition (Beck, 2010) but not vocal acoustics. While the majority of chimpanzees at Chimfunshi are rescued from adverse circumstances such as the pet trade, all subjects in the present study were born in captivity. However, the majority of their mothers were originally wild. Each subject belongs to one of four mixed-sex groups that comprise between 10 – 52 members, including infants, juveniles, sub-adults, and adults. For a further breakdown of the study population as a function of age, sex, and group, see table 2.1.

Table 2.1 Overview of subject characteristics including developmental stage, age, sex, and group membership.

Subject	Developmental stage	Age	Sex	Group
Bill	Infant	2 months	Male	3
Joey	Infant	11 months	Male	2
Rogers	Infant	11 months	Male	1
Victoria	Infant	5 months	Female	2
Murial	Infant	10 months	Female	2
Nancy	Infant	1 year	Female	2
Dennis	Infant	2 years	Male	2
Don	Infant	2 years	Male	2
Mumba	Infant	2 years	Male	2
Gloria	Infant	2 years	Female	2
Ida	Infant	2 years	Female	1
Tom	Infant	3 years	Male	2
Tina	Infant	3 years	Female	2
Brent	Juvenile	4 years	Male	3
Chitalu	Juvenile	4 years	Female	2
Jewel	Juvenile	5 years	Male	4
May	Juvenile	5 years	Female	2
Danny	Juvenile	6 years	Male	2
Little Jack	Juvenile	6 years	Male	2
Martin	Juvenile	6 years	Male	2
Ireen	Juvenile	6 years	Female	1

Jones	Juvenile	7 years	Male	2
Kenny	Juvenile	7 years	Male	4
Lods	Juvenile	8 years	Female	3
Bruce	Juvenile	10 years	Male	3
Jack	Juvenile	10 years	Male	4

The four groups lived in four outdoor enclosures, respectively. The miombo woodland is their habitat and that of many wild chimpanzee populations (Schoeninger, Moore, & Sept, 1999). Here, chimpanzees are fed once daily at approximately 12pm, and some chimpanzees are fed indoors. Beyond this, chimpanzees at Chimfunshi may forage on naturally fruiting trees. The enclosures and group sizes are large enough that fission fusion social dynamics take place as well as regular territory patrols. At night, all chimpanzees sleep outside and many show nesting behaviours.

Data collection

Audio recordings were collected between 7am and 6pm from June 2018 to October 2018 (excluding 12pm-1pm when daily feeding occurred) using a Sennheiser ME66 directional microphone. Recordings were collected only when the subjects were outdoors and the recordist was within 2–10 meters of the subject. The main approach in collecting recordings was to use a 5-minute focal sampling method following a randomized order each day that allowed us to have an equal and unbiased representation of the sample in this study. However, due to the large size of the enclosures and the dense forest inside them, subjects were often not visible. For efficiency, we therefore decided to wait for 5 minutes to observe a subject. If the subject was not visible, we then recorded the visible

subject who was next highest on the focal list. It was attempted to obtain two focal recordings on a subject in a single day – one in the morning (before 12pm) and one in the afternoon (after 1pm). 650 focal recordings were collected overall. Some recordings were incomplete as the subjects left the view of the camera for more than 30 seconds, meaning they could no longer be seen and therefore could not be identified as the potential caller. There were 44 incomplete focal recordings. Furthermore, when there were no visible subjects where 5-minute focal recordings could be taken for that day, the subjects were also recorded ad libitum. 79 ad libitum recordings were taken. These additional recordings were also included in the analysis of this study in order to increase the overall number of calls. All data collection was permitted by the University of Portsmouth Animal Welfare and Ethical Review Body (AWERB) and Chimfunshi Research Advisory Board (CRAB) (see appendices 4 & 5).

Overall, between 15 and 51 focal recordings were collected per subject. Focal observation time ranged between 1.24 and 4.25 hours per subject (Infants: $M=2.72 \pm SD=0.96$; Juveniles: $M=3.43 \pm SD=0.04$). The total duration of ad libitum recordings per subject ranged between 0.03 hours and 3.52 hours (Infants: $M=0.81 \pm SD=0.79$; Juveniles: $M=0.18 \pm SD=0.15$). The total duration of incomplete focal recordings ranged between 0.20 hours and 1.27 hours per subject (Infants: $M=0.70 \pm SD=0.33$; Juveniles: $M=0.55 \pm SD=0.28$). Overall, total observation time ranged between 1.73 hours and 5.45 hours (Infants: $M=4.19 \pm SD=1.31$; Juveniles: $M=4.16 \pm SD=0.53$) per subject.

Identifying calls

The coded unit of vocal behaviour was the call type, which is a broad category of calls (i.e. grunts) which contains distinct variants (i.e. food grunt, pant grunt, etc). Call types were chosen because there is wide agreement regarding the call types produced by young chimpanzees (Goodall, 1990; Plooi, 1984; Slocombe & Zuberbuhler, 2010) but whether young chimpanzees exhibit distinct subtypes (i.e. food grunts, pant grunts, etc) is currently unclear due to a lack of systematic study. Calls could be comprised of a single call element, or a series of call elements otherwise known as a call 'bout'. 768 calls were identified in total. The call types included grunts (N=382), whimpers (N=147), laughter (N=139), screams (N=41), hoo calls (N=41), barks (N=8), squeaks (N=6), and pant hoots (N=4). Calls were identified based on auditory cues followed by systematic visual inspection of spectrograms according to the definitions in table 2.2. The definitions were chosen based on similarities in the definitions of these call types by both pioneering and contemporary primatologists who studied both infants and adults (e.g. Goodall, 1990; Plooi, 1984; Riede, Owren & Arcadi, 2004; Kojima, 2008; Slocombe & Zuberbuhler, 2010). It is important to note that while pant hoots were identified, they were not analysed because they are calls made up from a series of different call types and are therefore not comparable to the other coded call types. Interestingly, only 4 pant hoots were observed in the entire observation period, and all were given by juveniles, whereas all other call types were found in both infants and juveniles. For a breakdown of the number of calls produced per subject see Appendix 1 supplementary methods. An inter-rater reliability test was performed on 20% of the total identified calls, and Cohen's Kappa revealed a good (See Cohen, 1960) level of reliability ($K=0.752$).

Table 2.2 Definitions of call types based on previous studies of mostly infant and adult chimpanzee vocalisations.

Call type	Definition	References
Grunt	Short, low-frequency calls given singularly or in short bouts. They may be tonal or noisy and produced with variable rhythm.	Plooij, 1984; Kojima, 2008; Slocombe & Zuberbühler, 2010
Whimper	Soft low-frequency tonal calls that can become higher in both frequency and amplitude as a bout progresses. Occasionally they may contain harmonics.	Plooij, 1984; Kojima, 2008; Slocombe & Zuberbühler, 2010
Laughter	Staccato, noisy, low-frequency, alternating ingressive-egressive breathing patterns delivered in an irregular rhythm. Acoustic energy is audibly present in both ingression and egression, with most energy visible during ingresses. While some adopt a more comprehensive laughter definition that includes grunt-like sounds (e.g. Davila-Ross et al. 2009), we decided to adopt a narrower definition with minimal overlap with other call types given the aim of this study.	Plooij, 1984; Slocombe & Zuberbühler 2010

Scream	Loud, high-frequency, harmonic vocalization with varying degrees of tonality (from calls with a high level of noise to calls with clearer harmonics and low noise levels). Non-linear phenomena are also commonly observed including frequency jumps (i.e. abrupt changes in fundamental frequency), subharmonics (i.e. spectral components at integer fractional values of the fundamental frequency and as harmonics of these values), biphonation (i.e. two simultaneous but independent fundamental frequencies), and deterministic chaos (i.e. episodes of non-random noise). Acoustic energy is usually present only during exhalation, but intense screams also include ingressive sounds.	Goodall, 1990; Riede et al. 2004; Slocombe & Zuberbühler, 2010
Hoo call	Tonal call with most energy at onset and a rise and fall in frequency over the call.	Goodall, 1990; Slocombe & Zuberbühler, 2010
Bark	Loud calls with abrupt onset. They are often noisy (either due to chaos or turbulence) with a low degree of tonality and are generally low pitched.	Goodall, 1990; Slocombe & Zuberbühler, 2010
Squeak	High-frequency, short calls often given in fast succession to form short bouts. These calls are predominantly clear tonal signals.	Plooij, 1984; Slocombe &

Zuberbühler,
2010

Pant hoot	A call series typically with four distinct phases. Firstly, an introductory phase of low frequency hoo calls. Second, a build-up phase consisting of increasingly loud panted hoo calls with energy visible on both ingression and egression. Third, a climax phase of screams. Finally, a let-down phase that resembles the build-up phase but with progressively decreasing energy.	Goodall, 1990; Slocombe & Zuberbühler, 2010
-----------	---	--

A call started with the onset of acoustic energy in the spectrogram and finished with the offset of acoustic energy in the spectrogram. A bout of elements of the same call type was coded as a single call, unless there was a gap between elements of 4 seconds or more. If another element occurred after 4 or more seconds, or if the call type changed, it was considered to be independent of the preceding element. This independence criteria was chosen for comparability with human infant research on this topic that employed a similar criteria (see Oller, 2000).

Acoustic analysis

Spectral and temporal acoustic features were extracted using the bioacoustics analysis programs Raven Pro V1.5 and Praat 6.0.43. Raven Pro was used for extracting all the acoustic data except for formants which cannot be measured using Raven Pro V1.5. In Raven, spectrograms were generated using a fast Fourier transform, whereas spectrograms in Praat were generated using linear predictive coding. Since the majority of extracted measurements were spectral sound characteristics rather than temporal,

narrowband spectrograms were chosen because these produce more accurate spectral measurements (Hopp & Owren, 2012). See table 2.3 for names of measured parameters and their definitions.

Table 2.3 Acoustic parameters and their definitions.

Parameter	Definition
Lowest frequency (Hz)	The lowest frequency at which energy was detected within the call.
Delta frequency (Hz)	The highest frequency at which energy was detected in the call minus the lowest frequency at which energy was detected in the call, corresponding to bandwidth.
Peak frequency (Hz)	The frequency where the highest amplitude value was observed.
Q1 frequency (Hz)	The frequency below which 25% of the total energy in the selection was found.
Q3 frequency (Hz)	The frequency below which 75% of the total energy in the selection was found.
Center frequency (Hz)	The frequency below which 50% of the total energy in the selection was found.
Mean entropy	Entropy measures the amount of disorder in the selection based on the distribution of energy within the selection. Mean entropy is calculated by first calculating entropy in each frame within the selection and then calculating the mean entropy across frames.

90% bandwidth (Hz)	The range of frequencies within which 90% of the total energy in the selection was found.
F1 (Hz)	Formants are resonant frequencies of the vocal tract where acoustic energy is concentrated. F1 refers to the lowest formant in the selection.
F2 (Hz)	The second lowest formant in the selection.
F3 (Hz)	The third lowest formant in the selection.
Duration (s)	The amount of time between the onset and offset of a call.
Element number	Number of call elements in a call.
Element rate (s^{-1})	The number of call elements in a call divided by the duration of that call.

A band-pass filter was applied to the spectrograms, ranging from 50 – 20,000 Hz. This bandwidth represented the range of frequencies where energy was visible in previous chimpanzee studies (Mitani, Gros-Louis & Macedonia, 1996; Slocombe & Zuberbühler, 2007). The sampling rate was 48,000 Hz with 16-bit accuracy. A Hanning window function was applied to call selections, which is the most appropriate window function for biological signal analysis because it prevents variation in onset-offset sound characteristics from introducing mathematical artefacts into the acoustic measurements (Clements, 1998). In Praat, identical settings were used as in Raven Pro. See figure 2.1 for an annotated spectrogram illustrating spectral measurements.

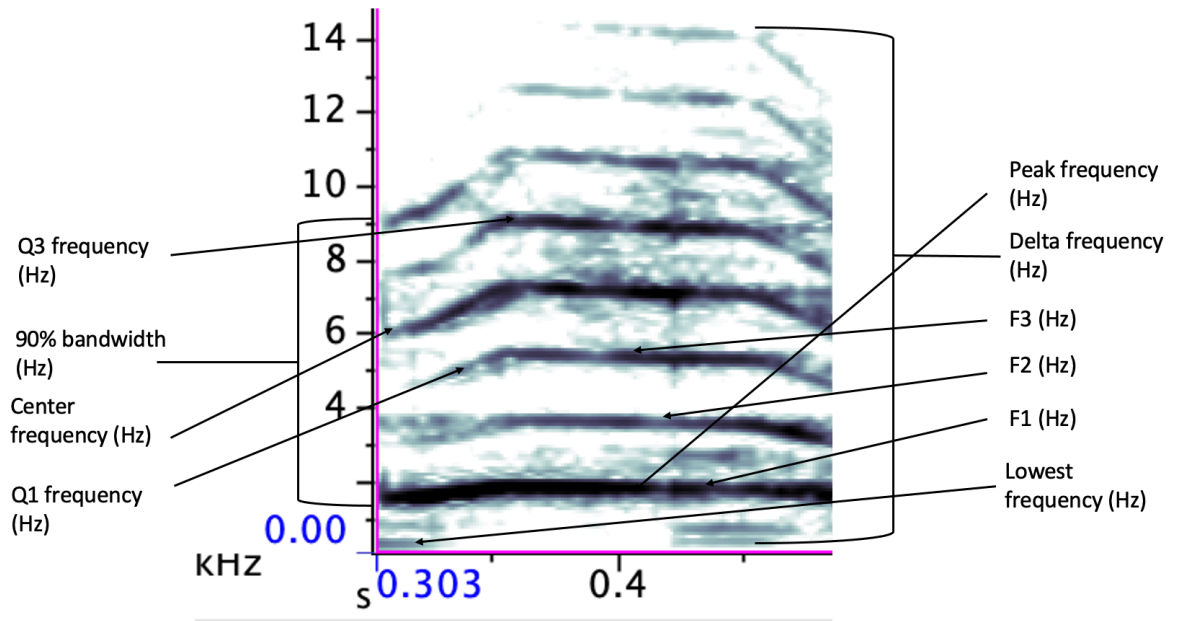


Figure 2.1 Annotated spectrogram illustrating spectral measurements.

Call elements were manually selected by highlighting the lowest frequency where there was observable acoustic energy, the highest frequency where there was observable energy, the onset of the call, and the offset of the call. We aimed to highlight the full call in a single selection. When ambient sound visibly overlapped with call elements of interest, we did not include these elements in the selection. Consequently, a single call could comprise several selections. Acoustic measurements were taken for each selection in the call series and the mean average of measurements was taken by dividing the sum total of those measurements by the number of selections taken. The mean value of measurements was not taken for lowest frequency, delta frequency, duration, element number, and element rate. For lowest frequency the lowest observed value across all selections was chosen. Delta frequency was calculated based on the lowest and highest observed frequencies across all selections. Duration was measured by highlighting the entire call series. Element number was measured by manually counting the number of

elements within the call and element rate was calculated by dividing the element number by the duration of the call. Since selections were manually taken by highlighting the lowest frequency where there was observable acoustic energy, the highest frequency where there was observable energy, the onset of the call, and the offset of the call, the only parameters that were likely impacted by the selection process were low frequency, delta frequency and duration. However, stability analyses (see statistical analysis section) were performed which measure whether a small number of parameters have a large impact on the model, which they did not appear to (see results section).

Statistical analysis

Fuzzy c-means clustering was applied to specify the number of clusters of call types in the data, to evaluate the validity of those clusters (how separable they are from one another), the reliability of those clusters (how consistently clusters could be extracted under varying degrees of overlap between clusters), the stability of those clusters (the extent to which cluster differentiation depends on a small number of variables), and to quantify the degree of overlap between clusters (as a measure of acoustic gradation).

Fuzzy c-means clustering algorithms cannot process datasets with missing data points. Consequently, we removed call cases where measurements could not be taken, or where some measurements were missing. This resulted in 447 vocalisations, including 221 grunts, 102 whimpers, 83 laughs, 28 screams, 5 barks, 4 squeaks, and 4 hoo calls. 259 of the calls were infant calls and 188 were juvenile calls. The composition of the infant and juvenile datasets regarding call types were very similar (see results section). Since many of the acoustic variables were measured on different scales (i.e. seconds, Hz, etc), before

any models were created, all variables were z-transformed to prevent the influence of a parameter on cluster solutions being based mostly on the range of the scale of the parameter, rather than systematic variation within that scale. Z-transformation in particular was chosen because it has been shown to lead to more accurate clustering solutions than other transformation techniques (Mohamad & Usman, 2013), and this technique has also been successfully applied to studies that aim to quantify acoustic call gradation in primates (Wadewitz et al. 2015; Fischer et al. 2017) and birds (Wonke & Wallschläger, 2009).

There are two key parameters that can be controlled in fuzzy c-means cluster analysis - K and μ . K specifies the number of clusters to be extracted. μ specifies the degree of overlap between clusters that is tolerated and is known as the ‘fuzziness parameter’. When $\mu = 1$, extracted clusters are not permitted to overlap. When $\mu > 1$, clusters are permitted to overlap. The greater the value of μ , the more clusters are permitted to overlap. μ can be continuously increased until cluster membership coefficients, which quantify the extent to which a given observation belongs to each cluster and ranges between 1 (complete membership) and 0 (not a member), reach $1/K$. At this point, cluster membership coefficients are equal across all clusters and the algorithm can therefore no longer assign cluster membership to observations (Zadeh, 2008; Zhou, Chao, & Yang, 2014).

Regarding the K values, we chose a minimum of 2 clusters in order to quantify gradation between clusters in the dataset which requires at least 2 clusters, and a maximum of 7 clusters because we included 7 call types in the data. Differences between call types in

the number of cases included (i.e. only 5 barks compared to 221 grunts) are not problematic in fuzzy c-means clustering wherein there is no minimum requirement for number of datapoints per cluster because clustering is based on *individual* acoustic characteristics of each case rather than the overall number of calls (Gamba et al., 2013; Wadewitz et al., 2015). μ values were systematically varied between 1.1 and 5, at increments of 0.5. This range and increment was chosen because the optimal description of both simulated and real-world data in terms of cluster validity is typically found in the range of $\mu=1.1-5$ (Zhou et al. 2014). Importantly, previous studies that applied fuzzy c-means clustering to acoustically graded primate vocal behaviour found the most stable solution for describing the data had $\mu=2$ (Wadewitz et al. 2015; Fischer et al. 2017), which is within the range of μ values used in this study.

Cluster validity & reliability

By systematically varying the K and μ values, 162 solutions were generated, 54 overall solutions, 54 infant solutions, and 54 juvenile solutions. Only infant and juvenile solutions are shown in the main text. All models converged within 500 iterations. The validity of each solution was evaluated based on the mean silhouette value for that solution. Silhouette values represent the degree of confidence that a data point belongs to its primary cluster, as measured by the ratio of the Euclidean distance between a data point and its nearest neighbour within its primary cluster, relative to the Euclidean distance between a data point and the nearest observation in the nearest neighbouring cluster (Bezdek, 1973). Silhouette values can vary between -1 and +1 with values > 0 indicating some degree of confidence regarding cluster membership (Bezdek, 1973). Each data-point has a corresponding silhouette value, therefore, mean silhouette values were used

to measure the confidence of the overall solution. Higher mean silhouette values result from clusters being more separable from one another.

The range of μ values over which solutions can be computed for any given K value indicates how reliable a model is (Fischer et al., 2017). The value of μ constrains the amount of overlap between clusters, with higher values allowing more overlap. Since the fuzzy c-means algorithm cannot compute a solution when cluster memberships for that solution are too close to $1/K$, if a solution can be extracted only for low values of μ , this suggests the model is not reliable (i.e. robust to overlap between clusters). Models were run and silhouette measurements were extracted in R using the 'fanny' implementation of the fuzzy c-means clustering algorithm from the 'cluster' package (version 1.15.3) (Maechler, 2019).

Cluster stability

Stability measurements compare the results of the fuzzy c-means clustering algorithm when one variable at a time is systematically removed. Using the 'clValid' function from the 'clValid' package in R (version 0.6.4), 4 stability measurements were calculated – mean proportion of non-overlap (APN), mean distance (AD), mean distance between means (ADM) and figure of merit (FOM). A range of measurements was chosen to capture different dimensions of stability, and compare not just overall stability between solutions, but ways in which different solutions are stable. APN measures the mean proportion of data points that change primary cluster membership when the model is systematically recalculated with one variable missing. APN values vary between 0 and 1 with lower values indicating higher stability. AD measures the mean Euclidean distance between

observations in the same cluster when the model is systematically recalculated with one variable missing. AD values vary between 0 and ∞ with lower values indicating higher stability. ADM measures the mean Euclidean distance between cluster centres and observations in the same cluster when the model is systematically recalculated with one variable missing. ADM values vary between 0 and ∞ with lower values indicating higher stability. Finally, FOM measures the mean intra-cluster variance of observations in the deleted variable, where clustering is based on the remaining samples. FOM values vary between 0 and ∞ with lower values indicating higher stability. See Brock et al. (2008) for further information on cluster stability measurement in clValid. Together, these analyses measure the extent to which extracted clusters rely on a small number of variables.

Cluster composition

After cluster validity and stability was assessed, a hard clustering solution was extracted for the best fitting overall, infant and juvenile models. In all cases a 2-cluster solution was the best fitting model and therefore all calls were allocated a value of either 1 or 2 depending on whether their primary cluster membership values were highest for cluster 1 or cluster 2. We then cross-tabulated hard cluster membership and call type to examine the contents of each cluster in terms of coded call types. This was done to gain information about the relationship between coded calls and the optimal description of the data based on acoustic information alone, as well as to evaluate the comparability of infant and juvenile call clusters.

Cluster typicality

‘Typicality’ coefficients were extracted, which can be used to quantify the degree of acoustic gradation (Wadewitz et al. 2015; Fischer et al. 2017). Typicality coefficients were calculated by subtracting secondary cluster membership values from primary cluster membership values. Membership values were obtained using the aforementioned ‘fanny’ implementation of the fuzzy c-means clustering algorithm in the ‘cluster’ R package. Typicality coefficients vary between 0 and 1, with values indicating the percentage of overlap between other clusters. For example, a typicality value of 0.6 indicates 60% of the acoustic space covered by the call overlaps with its primary cluster. Therefore, lower typicality coefficients indicate a higher degree of gradation because this shows more of the acoustic space covered by a call is shared between clusters.

Typicality coefficients were extracted for the best fitting (defined as the model with the highest silhouette value) overall, infant, and juvenile models, as well as their counterpart models (i.e. if the best fitting infant and juvenile models had different parameters, infant and juvenile models were created with their own optimal parameters and the optimal parameters for the other data set) in order for values to be comparable across models. For instance, the degree of gradation is highly constrained by the value of the μ parameter because this parameter specifies the amount of overlap between clusters that is tolerated, meaning when $\mu = 1.1$ many more observations would have 100% primary cluster membership compared to when $\mu > 1.1$. Therefore, to make comparisons of typicality coefficients across models, those models must have the same μ values. For each model, typicality was calculated for the whole model to measure the overall amount of gradation, as well as for each cluster, to measure the extent to which clusters were distinct. Typicality coefficients were not normally distributed in any case and were

therefore directly compared between datasets using the Mann-Whitney U test, to test whether there were significant differences in the degree of acoustic gradation in infant vocalisations compared to juveniles. For further information on typicality measurements, see Fischer et al (2017).

The distribution of typical vocalisations was assessed using the approach outlined by Wadewitz et al. (2015). The halved mean absolute deviation of typicality coefficients was calculated and calls with typicality coefficients greater than this value were classified as 'typical', while calls with typicality coefficients lower than this value were classified as 'atypical'. This was performed on all models for which typicality coefficients were calculated. This provided a measurement of how many calls are typical of calls in their own cluster, to complement the raw typicality coefficients which quantify how typical those calls were. These complimentary measurements are only reported in the supplementary results sections c-f (appendix 1).

Results

Cluster validity, reliability, and stability

For both infants and juveniles, 2-7 clusters were successfully identified and 2-cluster solutions consistently showed the highest silhouette value (figure 2.2). The μ value with the highest silhouette value for infants was 1.5 which generated a mean silhouette value of 0.382 (figure 2.2), whereas for juveniles the highest silhouette value was obtained when $\mu=1.1$, with a mean of 0.304 (figure 2.2). This suggests that there is greater overlap between clusters in the juveniles dataset compared to the infant dataset because silhouette values represent the separability of clusters and μ constrains overlap between

clusters. Therefore, if the highest silhouette values occur when the μ value is low this suggests allowing overlap between clusters reduces the separability of clusters, potentially indicating more graded call acoustics.

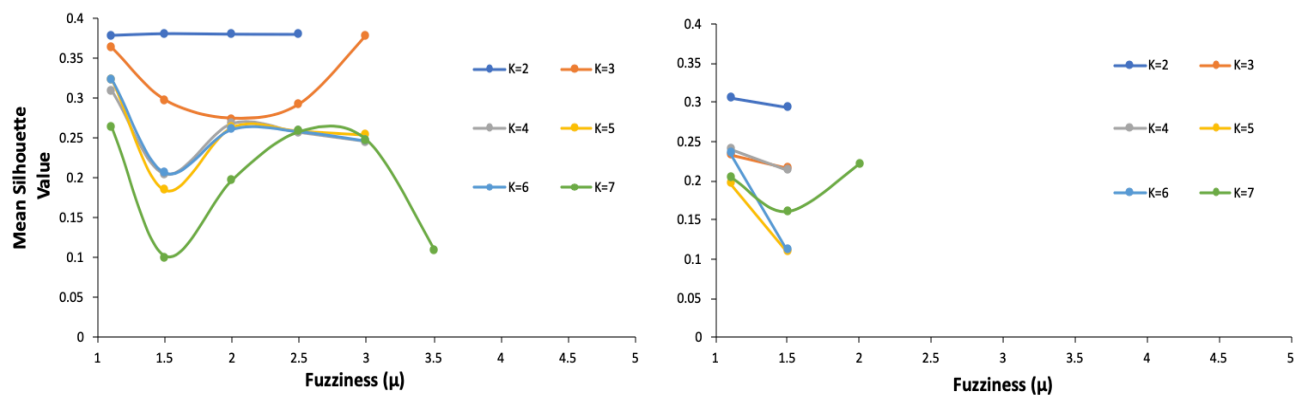


Figure 2.2 Line graph representing the mean silhouette values as a function of μ and K for the infant (left) and juvenile (right) datasets.

For infants, 2-cluster solutions could be calculated from $\mu=1.1$ to $\mu=2.5$. However, all other solutions could be calculated up to $\mu=3.0$, except the 7-cluster solution which could be calculated up to $\mu=3.5$ (figure 2.2). By contrast, juvenile clusters could only be extracted up to $\mu=1.5$ with the exception of K=7, $\mu=2.0$. The difference between infants and juveniles in the range of μ values over which clusters could be extracted implies that juvenile vocal behaviour is more graded than infant vocal behaviour, because poor model performance at higher values of μ results from a high degree of overlap between clusters. That 7-cluster solutions could be calculated at a larger range of μ values suggests that 7 clusters could be a more reliable description of the data for both infants and juveniles. However, 7 cluster solutions for infants and juveniles had consistently lower silhouette

values than 2-cluster solutions and were also less consistent in the silhouette values they generated, which varied from 0.101-0.258 for infants (2-cluster solution range = 0.378-0.381) and from 0.160-0.221 for juveniles (2-cluster solution range = 0.292-0.303). This suggests that, despite being more reliably extracted, 7-cluster solutions are less consistent in the confidence of call membership per cluster compared to 2-cluster solutions.

In comparison to the infant data set, the mean silhouette values for juveniles cluster solutions where $K > 2$ did not come as close to the values obtained for models where $K=2$. The mean silhouette value for the next best performing model with $K > 2$ for juveniles ($K=3$, $\mu=1.1$) was 0.075 lower than the best performing model. For infants at $\mu=1.1$, the silhouette value for the 3-cluster solution was only 0.01 lower than its counterpart 2-cluster model. Furthermore, at $\mu=3.0$, a 2-cluster solution for infants could not be extracted but a 3-cluster solution whose silhouette value was only 0.003 lower than the best performing model was found. A comparably well-fitting model with > 2 clusters was not observed for juveniles. This suggests that not only were infant clusters more discrete, but that there was little evidence in the juvenile dataset that there could be a higher number of discrete clusters depending on the μ values, whereas there was evidence for a third discrete cluster when $\mu=1.1$ and $\mu=3$ in the infant data.

Importantly, while infant and juvenile models differ in both the validity and reliability of clusters, both infant and juvenile 2-cluster solutions were highly stable. When the models were systematically recalculated with 1 parameter missing each time, APN values showed a mean of only 2% of datapoints changed cluster membership for infants and a mean of

4.5% of datapoints changed cluster membership for juveniles (see appendix 1, supplementary results section a for further measurements). This shows that 2-cluster distinctions were not heavily reliant on a small number of acoustic parameters.

Cluster composition

In the best fitting infant model ($K=2$, $\mu=1.5$), cluster 1 predominantly consisted of grunts, while cluster 2 predominantly consisted of whimpers, but also contained a high number of grunts. The majority of all call types except whimpers and hoo calls belonged to cluster 1 (table 2.4). Cluster compositions for the best fitting juvenile model ($K=2$, $\mu=1.1$) were very similar to that observed in the best fitting model for infants (table 2.4). Cluster 1 predominantly consisted of grunts, while cluster 2 predominantly consisted of whimpers, but also contained a high number of grunts. Cluster compositions were highly similar for both infants and juveniles with the exception of squeaks, which were found only in cluster 2 for the juvenile model, while they were found mostly in cluster 1 in the infant model. However, since only 1 squeak was observed in the juvenile dataset, this is unlikely to indicate any kind of important shift of juvenile vocal behaviour. Further, the difference in cluster composition is unlikely to confound infant-juvenile comparisons because these comparisons are made at the level of the cluster rather than the call type. Cluster composition for the counterpart (non-best fitting) infant and juvenile models are highly similar to those shown here (supplementary results section b, appendix 1). Cluster compositions for the overall model $K=7$, $\mu=1.1$ demonstrating validity of original call coding is shown in supplementary results section c (appendix 1). Descriptive statistics on the acoustic characteristics of all observed call types are shown in supplementary results section h, and descriptive statistics on the acoustic characteristics of the best-fitting

infant and juvenile models and their counterpart models are shown in supplementary results section I (appendix 1). This generally shows cluster 1 is comprised of wide-band calls with energy focused at higher frequencies while cluster 2 is comprised of more low-frequency tonal calls.

Table 2.4 Total number of calls per call type per cluster and percentage of each call type per cluster for infants and juveniles respectively.

Cluster	<u>Call Types</u>						
	Grunt	Whimper	Laugh	Scream	Hoo	Bark	Squeak
Number							
1							
Infants	103	12	30	11	1	3	2
	(70.55%)	(17.39%)	(93.75%)	(100%)	(14.29%)	(75.00%)	(66.66%)
Juveniles	52	7	40	17	0	1	0
	(69.33%)	(21.21%)	(78.43)	(94.44%)	(0%)	(100%)	(0%)
2							
Infants	43	57	2	0	6	1	1
	(29.45%)	(82.61%)	(6.25%)	(0%)	(85.71%)	(25.00%)	(33.33%)
Juveniles	23	26	11	1	8	0	1
	(30.66%)	(78.29%)	(21.57%)	(5.56%)	(100%)	(0%)	(100%)

Cluster typicality

Typicality coefficient values are highly constrained by the value of the fuzziness parameter because typicality reflects the degree of overlap between clusters and the

degree of cluster overlap is constrained by the value of the fuzziness parameter. Thus, to compare typicality across datasets, typicality was calculated for both the best-fitting model and the corresponding model in another dataset, even if that model was not the best fitting model. It is important to note that there was very little difference in cluster validity between the best fitting model and the alternative models selected for direct comparison (see figure 2.2). As such, the data were comparable within and between ontogenetic stages (i.e. infants and juveniles).

In the best fitting infant model ($K=2$, $\mu=1.5$), mean typicality was relatively high (0.65), ranging between 0.48 and 0.82 across subjects (figure 2.3). This pattern was consistent across call types, where mean typicality ranged between 0.62 and 0.81 (see appendix 1, supplementary results section d). In the corresponding juvenile model, mean typicality was relatively low (0.47) and coefficients ranged from 0.41-0.70 across subjects (figure 2.3), suggesting higher acoustic gradation in juvenile vocalisations. Mean typicality for juveniles was mostly low across call types, ranging between 0.37-0.71 (see appendix 1, supplementary results section d), suggesting higher acoustic gradation among juveniles is not attributable to changes in a single call type. The suggestion of increased acoustic gradation in juvenile vocalisations was supported by a Mann-Whitney U test comparing the mean typicality per subject between infants and juveniles when $K=2$, $\mu=1.5$ which showed that juvenile typicality measurements were significantly lower than infant measurements ($U=182$, $P=0.0002$). For a further breakdown of infant and juvenile typicality per call type per cluster for the best-fitting infant model and its juvenile counterpart, see appendix 1 supplementary results section e.

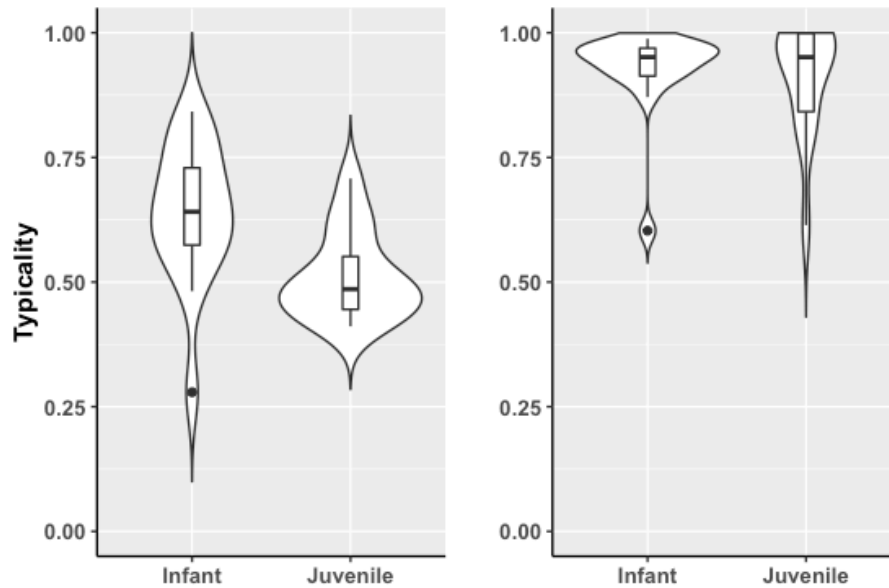


Figure 2.3 Violin plots depicting the medians, 25th percentiles, 75th percentiles, lower adjacent values, upper adjacent values, and probability density of mean typicality values per individual for infants and juveniles when $K=2$, $\mu=1.5$ (left) and when $K=2$, $\mu=1.1$ (right).

In the best fitting juvenile model ($K=2$, $\mu=1.1$), mean typicality was high (0.91) and coefficients ranged between 0.61 and 0.99 across subjects (figure 2.3). This pattern was consistent across call types, with mean typicality coefficients across call types ranging between 0.82 and 0.99 (see appendix 1, supplementary results section f). In the corresponding infant model, mean call typicality was also relatively high (0.95), ranging from 0.87-0.98 across subjects (figure 2.3), suggesting similar mean call typicality between infants and juveniles in these models but a wider range of observed values in juveniles. Infant mean call typicality was also consistently high across call types, ranging between 0.89-0.99 (see appendix 1, supplementary results section f). The apparent similarity between infant and juvenile call typicality in these models was supported by a direct comparison between infant and juvenile typicality when $K=2$, $\mu=1.1$ using a Mann-

Whitney U test which showed that the differences in typicality between these models were not significant ($U=116$, $P=0.619$). Spectrograms of the calls with the highest typicality values for cluster within each model are shown below in figure 2.4. For a further breakdown of infant and juvenile typicality per call type per cluster, see appendix 1 supplementary results section g.

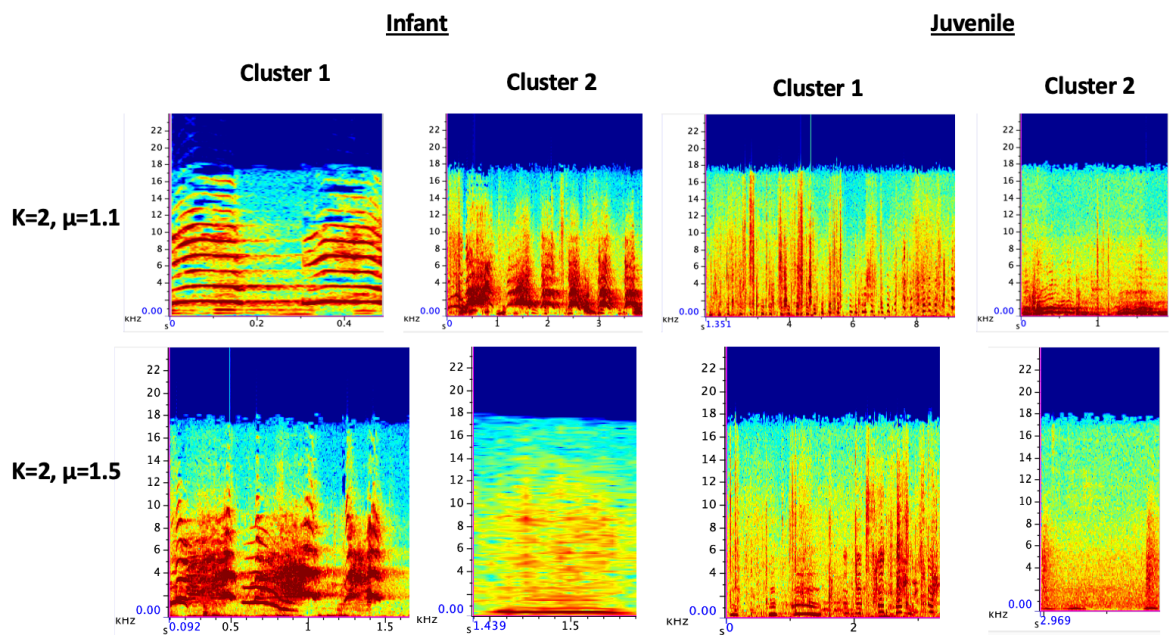


Figure 2.4 Spectrograms of the calls that generated the highest typicality values for each cluster in each best fitting infant and juvenile model and their counterpart models.

Discussion

In the present study, we aimed to examine ontogenetic changes in repertoire size and acoustic gradation in order to evaluate whether vocal type expansion occurs during chimpanzee vocal ontogeny. Fuzzy c-means clustering was applied to an acoustic dataset derived from the vocalisations of chimpanzees aged 0–10 years. Infant and juvenile repertoires were best described using the same number of clusters, suggesting the vocal

repertoire does not become more diverse with respect to the number of discrete call clusters throughout the first 10 years of chimpanzee vocal ontogeny. To be clear, any given set of clusters that is extracted is only one possible classification. Thus, while we did not find any evidence for an increase in repertoire size, we do not contend that the extracted solutions necessarily correspond to the true vocal repertoire, nor that this analysis contains all possible repertoire sizes. However, infant call clusters were more acoustically distinct from one another, and were apparently more reliable as the algorithm was able to identify infant clusters over a larger range of fuzziness values compared to juvenile call clusters. This inference was supported by analysis of cluster ‘typicality’, which showed that infant call typicality was significantly higher than juvenile call typicality, indicating infant vocalisations are significantly less acoustically graded than juvenile vocalisations.

The observed differences in infant and juvenile call acoustics could be partially attributable to differences in morphology of the vocal apparatus. MRI studies have shown that in chimpanzee ontogeny, much like humans, both the larynx and the hyoid bone descend deeper into the throat, although not to the same extent as is observed in humans (Nishimura et al., 2003, 2006). This opens up a wider acoustic space of possible sounds that can be produced (Lieberman, 1984; Crelin, 1987), meaning these morphological changes may represent a physiological mechanism by which the acoustic space between call clusters becomes progressively utilized in chimpanzee ontogeny. However, more recent studies of primate vocal production have shown that even with a larynx positioned high in the throat, an acoustic space of sounds comparable to that of humans is available, leading researchers to argue changes in the diversity of sound

production are better explained by neurological changes that underlie vocal control rather than vocal tract morphology (Fitch et al., 2016).

Morphological and neurological changes may provide proximate explanations of ontogenetic changes, however, they do not fully address the ultimate question of why call types are more acoustically distinct earlier in ontogeny and more acoustically graded later in ontogeny. Signals that are more acoustically consistent are theorised to be more easily interpreted by receivers (McCowan, Hanser & Doyle, 1999). Since the developmental timing of phenotype emergence can be selected to allow individuals to navigate different ontogenetic niches (Werner & Gilliam, 1984), perhaps less acoustic gradation in the chimpanzee infant vocal repertoire may be a trait that has been selected to help infants navigate their early ontogenetic niche wherein individuals are heavily dependent on caregivers for survival (Plooij, 1984), by reducing ambiguity in signals. For juveniles, who have increasingly complex social lives involving non-kin group members (Goodall, 1990), an increased acoustic space of sounds may offer valuable new communicative opportunities, which can be important for navigating more complex social niches (Freeberg et al., 2012).

While we observed increased acoustic gradation in juvenile vocal behaviour, which may be related to changes in the chimpanzee ontogenetic niche, it is important to consider that the juvenile social world also differs from the adult chimpanzee social world. For example, adult males will typically join a core group of dominants and remain within their natal community, while females will often migrate to new communities and spend much more time in small family units (Goodall, 1990). Since chimpanzee communities are

known to differ in the acoustic structure of their pant-hoot calls which are used in between-group interactions (Mitani et al., 1992; Marshall et al., 1999), female migration may require vocal flexibility to be able to produce a local pant-hoot variant.

Consequently, to more fully understand chimpanzee vocal ontogeny, it will be essential to quantify repertoire changes later into adulthood. This will help to better understand the social significance of increased acoustic gradation in juvenile vocal behaviour relative to infants, and the role that this plays in subsequent vocal ontogeny. More discrete vocal categories suited to the adult ontogenetic niche may emerge from highly acoustically graded juvenile vocal behaviour. Since the adult niches involve different social challenges for males and female, different patterns of vocal ontogeny may also be observed between sexes.

The bioacoustic complexity of signals as measured by diversity in acoustic form may represent a challenge for receivers to contend with because signals with an inconsistent acoustic form are expected to be more difficult to use for communication among living things (McCowan et al. 1999). As such, increased acoustic gradation in the repertoire may make signals more challenging for receivers to interpret. Comprehending acoustically graded signals may therefore require complex cognitive abilities, and socio-cognitive development in chimpanzees may be related to the increased use of acoustically graded signals. For example, in baboon ontogeny individuals improve in their ability to discriminate between acoustically graded bark variants and respond appropriately to them (Fischer et al. 2000). This example demonstrates that it is important to consider the extent to which vocal behaviour is perceived as graded also. Graded primate vocalisations have indeed been demonstrated to be perceived categorically (Fischer, 1998). While 2-

cluster solutions were most valid in the present study, solutions comprising 2-7 clusters for both infants and juveniles that could also be extracted, with some of those solutions outperforming 2-cluster solutions with regard to reliability. Neither the most valid nor the most reliable solution is necessarily the most meaningful set of call clusters for the chimpanzee because this depends not only on the acoustic properties of sounds, but how receivers attend to and react to different aspects of those signals. The present study simply identifies a set of possible clusters based on call acoustics and describes the gradation between them. This provides a starting point for subsequent studies to examine which description appears to be most meaningful to receivers based on their responses to calls from different clusters in different models.

The observed pattern stands in contrast to acoustic changes observed early in human ontogeny. Prior to the onset of canonical babbling, human infant vocal behaviour is highly acoustically variable and inconsistent, while later on in the first year of life, between 9-12 months of age, the repertoire expands into a larger number of more acoustically distinct vocal categories (Goldstein & Schwade, 2008; Esling, 2012; Mitchell and Kent, 1990). This pattern is predicted by Oller's (2012) infrastructural natural logic model, which posits that more discrete vocal categories emerge from vocal behaviour that is originally more acoustically graded. While there is evidence in some cercopithecine (Gouzoules & Gouzoules, 1989; Seyfarth & Cheney, 1986) and catarrhine monkey species (Snowdon, 1988) that vocal categories are less acoustically distinct in infancy and become more acoustically distinct later in ontogeny, our observations of chimpanzee vocal acoustics do not follow this pattern. In fact, we have observed the inverse pattern – vocal classes that are originally more distinct early in ontogeny, give rise to vocal behaviour that is more

acoustically variable later in ontogeny. Although comparisons between human and chimpanzee communication must be made with caution (see Rendell & Owren, 2013), we might also question the extent to which human speech is comprised of discrete units. Formal linguistics views language as a system with a finite number of discrete elements that can be recursively combined infinitely (Nishimura et al. 2003, 2006). However, when viewed phonetically, natural speech exhibits much acoustic gradation. For example, speech sounds are imbued with prosodic features – acoustic parameters that continuously vary and help to disambiguate lexical meaning by providing information about emotions and arousal (Hammerschmidt & Jürgens, 2007). Additionally, in some languages and dialects, both vowel (e.g. Hickey, 1984) and consonant (e.g. Ulving, 1953; Langacker, 1976) sounds have intermediate forms that grade into one another. Consequently, the degree of acoustic gradation in human vocal communication may have been underestimated, which could be explored by applying the present protocol to datasets on human vocal acoustics.

Given the aforementioned implications of the present findings, it is necessary to turn our attention to possible methodological issues in the present study. One might ask whether our finding of a lower degree of call gradation in the infant chimpanzees may be the result of the infants showing a propensity to produce call types which have a low degree of gradation more often, instead of having an overall repertoire that is less acoustically graded than the juvenile call repertoire. However, our examination of cluster composition with regard to call type showed that cluster compositions were very similar for infants and juveniles. Cluster 1 for both infants and juveniles comprised most of the grunts, screams, and laughs, while cluster 2 for both infants and juveniles comprised most of the

whimpers and hoo calls. The proportions of these calls that represented each cluster were also highly similar between infants and juveniles. Therefore, differences in the acoustic gradation of infant and juvenile vocalisations are not explained by a propensity to produce specific call types. Moreover, the observed differences between infants and juveniles in acoustic gradation cannot be explained by changes in a single call type (e.g. just grunts) rather than changes at the level of the repertoire as a whole, because typicality measurements were consistently lower for juveniles compared to infants across call types. Consequently, it appears that the difference in acoustic gradation between infants and juveniles reflects an ontogenetic shift in the acoustic characteristics of the entire repertoire.

One may also ask whether juvenile call typicality was lower because there were fewer calls in the juvenile dataset. In fact, we argue that a dataset with fewer examples would be expected to appear more discrete because it is less likely to capture the full extent of variation acoustic variation in calls. This is evident when one considers the process of cluster extraction: First, a centroid (the centre point of a cluster) is chosen in an n -dimensional space (where n = the number of acoustic variables). Next, a new data point is added, and the centroid is recalculated by averaging the data points in that cluster. Clusters are separated by finding lines in n -dimensional space that separate data into the most homogeneous clusters. As the centroids are iteratively recalculated, so too are the n -dimensional lines that separate clusters. When there are fewer data points, it is easier for the algorithm to find a line that separates those data points into different groups (Bezdek, 1973; Dunn, 1973). Consequently, when there is a smaller number of data points, we argue typicality should also be higher because typicality measures how typical

a call is of its respective cluster based on the extent to which it overlaps with other clusters, and when there are a small number of data points, overlap between clusters is less likely. Yet, the opposite pattern was found in the smaller dataset in this study, which suggests that acoustic gradation in the juvenile dataset is more likely to be underestimated rather than overestimated because if the juvenile dataset was larger, more overlap between clusters would be expected.

To conclude, in the present study we aimed to quantify ontogenetic changes in repertoire size and acoustic gradation in chimpanzee vocalisations. The results showed that between the infant and juvenile period, the number of call types observed did not differ, but juvenile call acoustics were significantly more graded than infants. This suggests that in contrast to human vocal ontogeny, chimpanzee vocal ontogeny is a process of filling in the acoustic gaps between early call types. Calls with greater acoustic diversity also have greater information encoding potential, raising the intriguing question of how chimpanzees make use of these new communicative possibilities during ontogeny.

3. Functional flexibility in chimpanzee vocal ontogeny

Abstract

From 3 months of age, human infants are able to express a full range of affective states with the same call type, achieving different responses from caregivers depending on what state the call expresses. This ability, known as 'functional flexibility', provides an essential ontogenetic foundation for language, but whether functional flexibility provides a phylogenetic foundation for language is less known. To gain insight into this question, we examined patterns of vocal production and function among infant (N=15) and juvenile (N=13) chimpanzees at Chimfunshi Wildlife Orphanage, Zambia. Following the approach used for human infants, we first classified subjects' affective states while they vocalised as either positive, neutral, or negative, based on facial and bodily cues. These vocalisations consisted of grunts, laughs, whimpers, screams, and hoo calls. To test for functional flexibility, we examined whether social partner responses to a call depended on what affective state that call expressed. We found that grunts showed no affective bias, whereas all other call types showed an affective bias towards either negative (whimpers, screams, and hoo calls) or positive (laughs) affective states. Grunts were also the only call type for which evidence of functional flexibility was observed, with positive grunts mostly eliciting behavioural continuation in a social partner, while neutral and negative grunts mostly elicited change. Notably, functional flexibility was mainly observed in juveniles. Overall, these patterns demonstrate that functional flexibility is present in the vocal behaviour of young chimpanzees as well as young humans, implying functional flexibility provides a foundation for the emergence of language in phylogeny as well as ontogeny.

Introduction

Early flexibility in human vocal production and function is theorised to be a foundation necessary for language to develop (Clay, Archbold, & Zuberbühler, 2015; Zuberbühler, 2011). Typically, signal and function are tightly coupled in animal calls (Oller, 2012). For example, acoustically distinct vervet alarm calls reliably elicit defensive behavioural responses from conspecifics that depend on the predator associated with the alarm (Seyfarth et al., 1980). By contrast, the same linguistic utterance can be flexibly produced to express different meanings which achieve different functions. As such, language is marked by a decoupling of signal and function. This ability to achieve different functions by expressing the same utterance in different ways is known as ‘functional flexibility’ (Oller et al., 2013).

Indeed, functional flexibility provides a foundation for language development in human ontogeny. In 3-12 month old human infants, Oller et al. (2013) found protophones, which are believed to be precursors to speech sounds (Jhang & Oller, 2017; Yoo et al., 2018), express positive, neutral, and negative affective states on different occasions, which was indicated by subjects’ facial expressions during vocal production. By contrast, non-speech related calls such as laughter and cries were biased towards positive and negative affective states respectively. This component of functional flexibility (i.e. flexible expression of affect with a single call type) is known as ‘free expressivity’ and is theorised to be necessary for functional flexibility but not equivalent to it (Oller, 2012). Functionally flexible calls fulfil different functions (i.e. effects on social partners) depending on how the call is expressed. For example, Oller et al. (2013) found that when human infant

protophones expressed positive affect, their caregivers mostly responded with attempts to ‘encourage’ continuation of the social interaction (i.e. turn-taking, continuing proto-conversation, imitation, and praise), whereas negative protophones were associated with attempts to ‘change’ the infants’ state (i.e. changing the situation, soothing, scolding, and distracting the infant).

It is important to note that contextually flexible vocal production is not synonymous with free vocal expressivity because the production of a call across different contexts does not necessarily mean the call *expresses* something different on different occasions. However, for a call to be freely expressed (and in turn functionally flexible), the caller must be able to produce it in more than a specific context because context limits what a call can express (Oller 2012; Oller et al., 2013). This is supported by the observation that human infant protophones are produced in a context-free manner before they are used to express different affective states and fulfil different functions on different occasions (Oller, 2012; Jhang & Oller, 2017). In turn, this suggests species that exhibit contextually flexible vocal production are potential candidate species for exhibiting free vocal expressivity and functional flexibility (Oller, 2012).

Many non-human primate species show contextually flexible vocal behaviour, including adult baboon grunts (Maciej et al., 2013; Rendall et al., 1999), adult vervet monkey alarm calls (Price et al., 2015), adult Japanese macaque coo call (Koda, 2004), adult and juvenile mangabey grunts (Range & Fischer, 2004), adult bonobo high hoots (Schamberg et al., 2016) and adult chimpanzee grunts (Slocombe & Zuberbuhler, 2010). Since some of these studies found calls produced across contexts show acoustic differences between contexts

(e.g. Koda, 2004; Maciej et al., 2013; Price et al., 2015; Schamberg et al., 2016), it could be argued that they are in fact not flexibly produced call-types but instead are distinct context-limited sub-types. However, work on human infants on this topic (e.g. Oller et al., 2013) was conducted at the level of the call-type rather than distinct sub-types, suggesting flexibility at the level of the call-type is most relevant for the study of free expressivity and functional flexibility. Notably, some non-human primate studies have also shown that different individuals use the same call-type for different functions (Chimpanzees: Hopkins et al., 2007; Orangutans: Lameira et al., 2013; Bonobos: Taglialatela et al., 2003). While this suggests some flexibility between individuals, this is not evidence of flexibility within individuals, which is a defining feature of functional flexibility (Oller, 2012; Oller et al., 2013). Thus, while there seems to be a plentiful supply of primate species that would be candidates for free vocal expressivity and functional flexibility, to date there are just two studies of non-human primates that have attempted to study this ability in a manner comparable to the human infant literature.

One study on adult bonobo peep calls (Clay, Archbold & Zuberbuhler, 2015), and another on infant chimpanzee grunts and whimpers (Dezecache et al., 2020), have claimed to find evidence suggestive of functional flexibility. In both studies, the authors classified subjects' affective states during vocalisations as either positive, neutral, or negative, based on their behaviours. Although only a small selection of behaviours was used to classify affective states which did not include facial expressions, as was used by Oller et al. (2013), both studies found calls (only infant grunts in the case of Dezecache et al., 2020) were produced across all affective 3 states. This provides evidence for free expressivity in chimpanzees and bonobos, although social partner responses were not assessed,

meaning these studies did not provide evidence of functional flexibility per se. This is an essential gap to fill because evolved signals must benefit signallers through their effects on others (Maynard-Smith & Harper, 2003; Maynard-Smith & Harper, 1995).

Furthermore, due to a focus on grunts and whimpers (e.g. Dezechache et al., 2020), it is not known how widespread this phenomenon is within the chimpanzee vocal repertoire. This is important to address in order to assess similarities between immature chimpanzee and human vocal behaviour in this regard, because several chimpanzee calls (i.e. laughs, screams and whimpers) are believed to be homologous with human calls (see Davila Ross et al., 2009; Schwartz et al., 2020) and may therefore be expected to show similar affective bias. Due to Dezechache et al.'s (2020) focus on infant chimpanzees in the first year, it is also unknown whether this capacity persists later into ontogeny as it does in humans, wherein this ability represents a crucial social competency that can be used to navigate complex social situations throughout the life-span (Oller et al., 2013).

In the present study, we aimed to study functional flexibility during chimpanzee vocal ontogeny, in a manner comparable to the human infant literature. We focused on semi-wild infant and juvenile chimpanzees from 0-10 years of age. We classified subjects' affective states as either positive, neutral or negative based on a combination of bodily cues and facial expressions that co-occurred with calls. We studied grunts, which have been shown previously to be freely expressed in infant chimpanzees by Dezechache et al. (2020), as well as whimpers, hoo calls, screams, and laughs, many of which are believed to be homologous with affectively biased human calls (see Davila Ross et al., 2009; Schwartz et al., 2020). For each call type, we tested for free expressivity by testing whether each call type was significantly biased towards expressing particular affective

states. We hypothesised that grunts will be freely expressed throughout ontogeny as previous studies showed infant grunts were freely expressed (Dezecache et al., 2020), whereas laughs, screams, whimpers and hoo calls will be affectively biased because they are mainly believed to be homologous with affectively biased human calls (Davila Ross et al., 2009; Schwartz et al., 2020; Dezecache et al., 2020). To examine functional flexibility, we examined whether social partners changed or continued their behaviours depending on the affective valence of a call, in line with functional flexibility research in human infants (see Oller et al., 2013). Since free expressivity is theorised to be an essential precondition for functional flexibility (Oller, 2012), we predicted grunts would be functionally flexible whereas other call types would not be, because grunts have been shown to be freely expressed (Dezecache et al., 2020) whereas whimpers, hoo calls, laughs, and screams are expected to be affectively biased (Davila Ross et al., 2009; Schwartz et al., 2020; Dezecache et al., 2020).

Method

Subjects and study site

Subjects were infant (N=15) and juvenile (N=13) semi-wild chimpanzees housed at Chimfunshi Wildlife Orphanage, Zambia. Infant ages ranged from 2 months up to 4 years of age ($M=1.76 \pm SD=1.14$). Individuals aged between 4 and 10 years were classified as juveniles ($M=6.87 \pm SD=1.65$). Subjects are known to belong to a mixture of sub-species including *Pan troglodytes troglodytes* and *Pan troglodytes schweinfurthii*. However, many sub-species memberships were not known, meaning it is possible that the sample contained members of other sub-species. All subjects were raised by their mothers during infancy. All infant subjects still lived with their biological mothers. Three juveniles did not

live with their mothers due to fatalities that occurred in years prior to the present study. All subjects in the present study were born and raised in semi-wild sanctuary conditions. Each subject belongs to one of four mixed-sex colonies that comprise between 10 – 52 members. For a further breakdown of the study sample as a function of age, sex and colony table 3.1.

Table 3.1 Overview of subject characteristics including developmental stage, age, sex, and colony membership.

Subject	Developmental stage	Age	Sex	Colony
Bill	Infant	2 months	Male	3
Renate's Baby	Infant	4 months	Male	1
Joey	Infant	11 months	Male	1
Rogers	Infant	11 months	Male	2
Victoria	Infant	5 months	Female	2
Murial	Infant	10 months	Female	2
Camilla	Infant	10 months	Female	2
Nancy	Infant	1 year	Female	2
Dennis	Infant	2 years	Male	2
Don	Infant	2 years	Male	2
Gloria	Infant	2 years	Female	2
Ida	Infant	2 years	Female	1
Mumba	Infant	2 years	Male	2

Tina	Infant	3 years	Female	2
Tom	Infant	3 years	Female	2
Brent	Juvenile	4 years	Male	3
Chitalu	Juvenile	4 years	Female	2
May	Juvenile	5 years	Female	2
Jewel	Juvenile	5 years	Male	4
Danny	Juvenile	6 years	Male	2
Ireen	Juvenile	6 years	Female	1
Little Jack	Juvenile	6 years	Male	2
Martin	Juvenile	6 years	Male	2
Jones	Juvenile	7 years	Male	2
Kenny	Juvenile	7 years	Male	4
Lods	Juvenile	8 years	Female	3
Bruce	Juvenile	10 years	Male	3
Jack	Juvenile	10 years	Male	4

The four colonies lived in four outdoor enclosures, respectively. The miombo woodland is their habitat and that of many wild chimpanzee populations (Schoeninger, Moore, & Sept, 1999). The enclosures range between 47 and 190 acres in size. Here, chimpanzees are fed once daily at approximately 12pm, and some chimpanzees are fed indoors. Beyond this, chimpanzees at Chimfunshi may forage on naturally fruiting trees. The enclosures and colony sizes are large enough that fission fusion social dynamics take place as well as regular territory patrols. At night, all chimpanzees sleep outside and many show nesting behaviours.

Data collection

Video and audio recordings were collected between 7am and 6pm from June 2018 to October 2018 (excluding 12pm-1pm when daily feeding occurred) using a Sony CX405 Handycam with a Sennheiser ME66 directional microphone attached. Recordings were collected only when the subjects were outdoors and the recordist was within 2–10 meters of the subject. The main approach in collecting recordings was to use a 5-minute focal sampling method that allowed us to have equal representation of the sample in this study. Focal recordings for each subject were collected in a new order each day determined by using a random number generator. However, due to the large size of the enclosures and the dense forest inside them, subjects were often not visible. For efficiency, we therefore decided to wait for 5 minutes to observe a subject. If the subject was not visible, we then recorded the visible subject who was next highest on the list. It was attempted to obtain two focal recordings on a subject in a single day – one in the morning (before 12pm) and one in the afternoon (after 1pm). 650 focal recordings were collected overall. In addition, there were also recordings that were incomplete as the subjects left the view of the camera for more than 30 seconds and could therefore no longer be seen and identified as the potential caller. There were 44 incomplete focal recordings. Furthermore, when there were no visible subjects for whom 5-minute focal recordings could be taken for that day, the subjects were also recorded *ad libitum*. 79 *ad libitum* recordings were taken. These additional recordings were also included in the analysis of this study in order to increase the overall number of calls. All data collection was permitted by the University of Portsmouth Animal Welfare and Ethical Review Body (AWERB) and Chimfunshi Research Advisory Board (CRAB) (see appendices 4 & 5).

We aimed to obtain 3 hours of recordings per subject to match human infant research on this topic (e.g. Oller et al., 2013). Overall, between 15 and 51 focal recordings were collected per subject, meaning focal observation time ranged between 1.24 and 4.25 hours of recording per subject (Infants: $M=2.72 \pm SD=0.96$; Juveniles: $M=3.43 \pm SD=0.04$). The total duration of ad libitum recordings per subject ranged between 0.03 hours and 3.52 hours (Infants: $M=0.81 \pm SD=0.79$; Juveniles: $M=0.18 \pm SD=0.15$). The total duration of incomplete focal recordings ranged between 0.20 hours and 1.27 hours per subject (Infants: $M=0.70 \pm SD=0.33$; Juveniles: $M=0.55 \pm SD=0.28$). Overall, total observation time ranged between 1.73 hours and 5.45 hours per subject (Infants: $M=4.19 \pm SD=1.31$; Juveniles: $M=4.16 \pm SD=0.53$).

Identifying calls

The coded unit of vocal behaviour was the call type, which is a broad category of calls (i.e. grunts) that contains distinct variants (i.e. food grunt, pant grunt, etc). Call types were chosen because there is wide agreement regarding the call types produced by immature chimpanzees (Goodall, 1990; Plooi, 1984; Slocombe & Zuberbühler, 2010) but whether immature chimpanzees exhibit distinct subtypes (i.e. food grunts, pant grunts, etc) is currently unclear due to a lack of systematic study. Moreover, in a recent study of infant and juvenile vocal behaviour (Taylor et al., under review), we demonstrated the validity of these call types by using unsupervised machine learning techniques to show that based on acoustic information alone and without call labels, a clustering algorithm grouped the majority of each call type into distinct call clusters, suggesting call types can be reliably distinguished from one another.

Calls could be comprised of a single call element, or a series of call elements otherwise known as a call 'bout'. 768 calls were identified in total. The call types included grunts (N=382), whimpers (N=147), laughter (N=139), screams (N=41), hoo calls (N=41), barks (N=8), squeaks (N=6), and pant hoots (N=4). Calls were identified based on auditory cues followed by systematic visual inspection of spectrograms according to the definitions used in Taylor et al. (under review) (see chapter 2). An inter-rater reliability test was performed on 20% of the total identified calls, and Cohen's Kappa revealed a good (See Cohen, 1960) level of reliability ($K=0.75$).

Coding and classifying subject behaviours of valence

When they were vocalizing, the subjects' affective state was classified as either positive, neutral, or negative. We used a combination of facial cues (in accordance with Oller et al., 2013) and bodily behavioural cues (in accordance with Dezechache et al., 2020). Oller et al. (2013) relied only on facial cues to classify human infant affective states but no primatological studies on this topic (e.g. Dezechache et al., 2020; Clay, Archbold & Zuberbuhler, 2015) included facial expressions in affect classification. Where possible, it is important to include facial expressions because they are some of the most extensively studied and reliable affective cues (e.g. Cecilione et al., 2017). However, facial expressions are not the only valid indicators of affect and are also often not visible in chimpanzees due to challenging observation conditions, hence we also utilised bodily behavioural cues. All cues were chosen based on previous studies that found a relationship between that cue and a particular affective state. Facial expression definitions were based on previous studies (see Parr, Waller & Vick, 2007; Bard, Gaspar & Vick, 2011). Bodily cue definitions

were used based on Plooij's (1984) ethogram. Affective state cues achieved good levels of inter-rater reliability ($K=0.73$). For facial expression types, bodily cues, definitions, and associated valence, see table 3.2.

Table 3.2 Definitions of facial expressions (based on the definitions of Parr, Waller & Vick, 2007 & Bard, Gaspar & Vick, 2011), bodily cues (based on the definitions of Plooij, 1984), and associated valence with supporting literature.

Behaviour	Definition	Valence
Facial		
Expressions		
Open-mouth bared-teeth	Cheeks and upper lips raised, lip corners pulled back, and lower lip depressed while the jaw is lowered so the mouth is open. Gums are usually displayed and both upper and lower teeth usually visible.	Negative (Parr, 2001; Bard, Gaspar & Vick, 2011)
Closed-mouth bared-teeth	Upper lip raised, lips corners pulled back, and lower lip depressed. Upper and lower teeth usually visible	Negative (Parr, 2001; Bard, Gaspar, & Vick, 2011)
Display		
Pout	Lips separated and funnelled outwards.	Negative (Goodall, 1990; Itakura, 1993; Fernández -Carriba, 2002)

Full Open-mouth Face	Jaw lowered so mouth is open, lips parted, and stretched horizontally. Bottom and top teeth visible.	Positive (Parr, 2001; Bard, Gaspar & Vick, 2011)
Half Open-mouth Face	Jaw lowered so mouth is open, lips parted, and stretched horizontally. Bottom teeth visible.	Positive (Parr, 2001; Bard, Gaspar, & Vick, 2011)
No Expression	Absence of any coherent facial muscle activations recognised as an expression in previous literature.	Neutral (Fernández-Carriba et al., 2002; Vick et al., 2007)

Bodily Cues

Play	Play wrestling, pirouetting, solitary play, and somersaulting (see Plooij, 1984)	Positive (Parr & Hopkins, 2000)
Grooming	Picking through the fur of another individual.	Positive (Keverne et al., 1989; Bard et al., 2014)
Breastfeeding	Having mothers nipple in mouth. Actual sucking movements may or may not be seen.	Positive (Bădescu et al., 2016a; Bădescu et al., 2016b)
Nuzzling	Unsuccessfully attempting to access the mothers nipple.	Negative (Goodall, 1990)
Aggression	Tantrums, hitting, slapping, pulling hair outside of playing, and dominance displays (see Plooij, 1984).	Negative (Goodall, 1990; Parr & Hopkins, 2000)

Self-scratching	Moving the nails over the skin of some part of the own body while bending the fingers.	Negative (Itakura, 1993; Baker & Aureli, 1997; Yamanashi, 2010)
Avoidance	Avoiding approaching conspecifics, resisting physical contact of conspecifics, and defending objects from conspecifics (see Plooi, 1984). All behaviours must occur outside of play.	Negative (Ferdowsian et al., 2011)
Other	All coded behaviours that were not included in the above were considered to be neutral. Common examples are locomotion, climbing, laying down, sitting, object manipulation, and travelling.	Neutral (Kano, Tanaka, & Tomonaga, 2008; Kano & Tomonaga, 2010)

If facial cues and bodily behavioural cues matched in their affective valence, then the affective state of the subject was classified as such. If one cue was either positive or negative and the other cue was neutral (e.g. negative facial cue and neutral bodily cue), then the affective state of the subject was classified by the former (e.g. negative state). This approach was adopted because cues that can indicate affective state are not necessarily expressed by both the face and the body simultaneously (e.g. Taglialatela et al., 2015; Wilke et al., 2017). Cases wherein the valence of the cues were contrasting (i.e. positive facial cue and negative bodily cue) were not included in the analysis because

there was no basis for deciding which affective state should be given priority. Such contrasting cues to affective valence occurred in fewer than 10% of cases.

Measuring social partner behaviour

To test for functional flexibility, social partner behaviour was also coded. The social partner referred to the individual that the subject was interacting with while vocalising. This was chosen based on who the subject was directing their behaviours towards and/or who was directing their behaviours towards the subject. In 59% of cases the mother was one of the social partners. In 55% of cases there was a social partner who was not the mother. During a given call bout, a subject could interact with multiple different social partners (hence the additive percentage of cases where the social partner was the mother or a non-mother conspecific does not equal 100%). In 9% of cases there was no social partner at all.

The behaviour of the social partner was coded 4 seconds before the subject's call, during the call, and 4 seconds after the call. This time frame was chosen because studies that evaluated functional flexibility in the vocal behaviour of human infants also examined caregivers behavioural responses to their infants calls within this time frame (Oller et al., 2013; Jhang & Oller, 2017). All observed behaviours were coded using the ethogram developed by Plooij (1984). Only behaviours directed towards the subject were included in the analysis. Social partner behaviour was intra-rater reliability tested using Cohen's Kappa which showed an excellent degree of reliability was achieved ($K = 0.88$). The original and reliability coding were conducted 9 months apart. See table 3.3 below for behaviour categories and the behaviours that comprised each category.

Table 3.3 Categories of social partner behaviour directed towards subject and behaviours that belong to each category. See Plooij (1984) for definitions.

Category	Behaviour directed to subject
Feeding offspring	Food sharing, breast feeding
Protecting	Defending, gathering subject
Comforting	Cradling, patting, embracing, bite-kissing, holding hand, extending hand
Playing	Play wrestling, tickling
Grooming	Grooming, inspecting
Approach	Approaching subject
Travel	Lowering back for subject to climb on, following, carrying subject
Avoidance	Avoiding, leaving, or breaking contact outside of play
Preventing breastfeeding	Covering nipple (mothers only)
Threat	Arm raising gesture, biting at (outside of play), dominance displays
Taking	Taking or pulling objects in subjects' possession
Causing discomfort	Pulling, slapping, hitting, dragging, pulling hair, pushing (outside of play).

Similar to Oller et al. (2013), we used this data to examine whether social partners continued or changed their behaviour. Continuing behaviour was defined as when the same category of behaviour was observed before, during, and after the subject vocalised. Behavioural change was defined as when a social partner either stopped a behaviour or started showing a new category of behaviour during or after the call was observed. Behavioural categories were not mutually exclusive (for example, a social partner could both approach and groom the subject during the vocalisation). If one behaviour was observed before, during, and after the vocalisation, but another started and or stopped, this was classified as behavioural change. Social partners were observed to continue or change their behaviour on 40.28% and 59.72% of occasions, respectively.

Statistical analysis

To test for free expressivity, we calculated the percentage of calls that occurred during a positive, neutral, or negative affective state for every subject. If a call type is freely expressed, it should not be biased towards expressing a particular type of affective state. To test whether grunts, laughs, screams, hoo calls, and whimpers are freely expressed, a Friedman test was used to test compare the percentage of calls across positive, neutral and negative affective states within-subjects. A significant difference would suggest that call type is affectively biased and is therefore less freely expressed than a call that is not significantly more likely to express positive, neutral, or negative affective states. This test was performed on each call type. Wilcoxon tests were used to examine which affective states calls were associated with where Friedman tests yielded significant results. Mann-Whitney U tests were also used to test for relative differences between infant and

juvenile calls in the percentage of calls that were positive, neutral, and negative for each call type (see Appendix 2 supplementary results A).

To test for functional flexibility, we calculated the percentage of calls within-subjects that occurred during a positive, neutral, or negative affective state when social partners responded with behavioural change. We also calculated the percentage of calls within-subjects that occurred during a positive, neutral, or negative affective state when social partners responded with behavioural continuation. Call types are considered functionally flexible if the way social partners respond to them depends on what affective state express (i.e. vocalising during positive affect is associated with social partners continuing their behaviour while vocalising during negative affect is associated with social partners changing their behaviour, see Oller et al., 2013). Friedman tests were used to compare the percentage of calls that occurred during positive, neutral, or negative affective states when social partners changed their behaviour and when they continued their behaviour. Wilcoxon tests were used to examine which affective states were mostly associated with a particular response type where Friedman tests yielded significant results. To test for relative differences in the relationship between affect and social partner responses, we compared the percentage positive, neutral and negative calls per call type that elicited change vs continuation in social partner behaviour within-subjects using Wilcoxon tests. For functional flexibility, only results for call types shown to be freely expressed are shown in the main text. Tests on affectively biased calls are shown in Appendix 2 supplementary results B. Direct infant-juvenile comparisons in the percentage of positive, neutral and negative calls per call type that elicited each social partner response were performed using Mann-Whitney U tests (see Appendix 2 supplementary results B).

All tests were performed on infants, juveniles and overall, using Hommel-Hochberg corrections for multiple testing. All p-values were asymptotic p-values from one-tailed tests. All analyses were conducted using IBM SPSS Statistics V27.

Results

For grunts, there was no significant difference in affective valence (Friedman test, overall: $\chi^2(3)=1.733$, $N=21$, $p=.210$; infants: $\chi^2(3)=1.686$, $N=12$, $p=.215$; juveniles: $\chi^2(3)=0.875$, $N=9$, $p=.323$), suggesting grunts are regularly produced during positive, neutral, and negative affective states (see figure 3.1A). See Appendix 2 supplementary results A for direct infant vs juvenile comparisons of affective state per call type. There was no significant difference in the affective valence of grunts associated with behavioural change in a social partner (Friedman test, overall: $\chi^2(3)=1.088$, $N=20$, $p=.290$; infants, $\chi^2(3)=.581$, $N=9$, $p=.374$ juveniles: $\chi^2(3)=3.000$, $N=11$, $p=.111$) (see figure 3.1F), inconsistent with functional flexibility. However, overall but not for infants or juveniles alone, there was a significant difference in the affective valence of grunts associated with behavioural continuation in a social partner (Friedman test, overall: $\chi^2(3)=7.283$, $N=17$, $p=.013$; infants: $\chi^2(3)=3.920$, $N=9$, $p=.071$; juveniles: $\chi^2(3)=3.714$, $N=8$, $p=.078$) (see figure 3.1F), suggesting some degree of functional flexibility. Post-hoc tests showed grunts associated with behavioural continuation in a social partner were more likely to be positive vs neutral (Wilcoxon test, $Z=-.2144$, $N=17$, $p=.032$) but not positive vs negative (Wilcoxon test, $Z=-1.972$, $N=17$, $p=.098$) or negative vs neutral (Wilcoxon test, $Z=.171$, $N=17$, $p=1.00$).

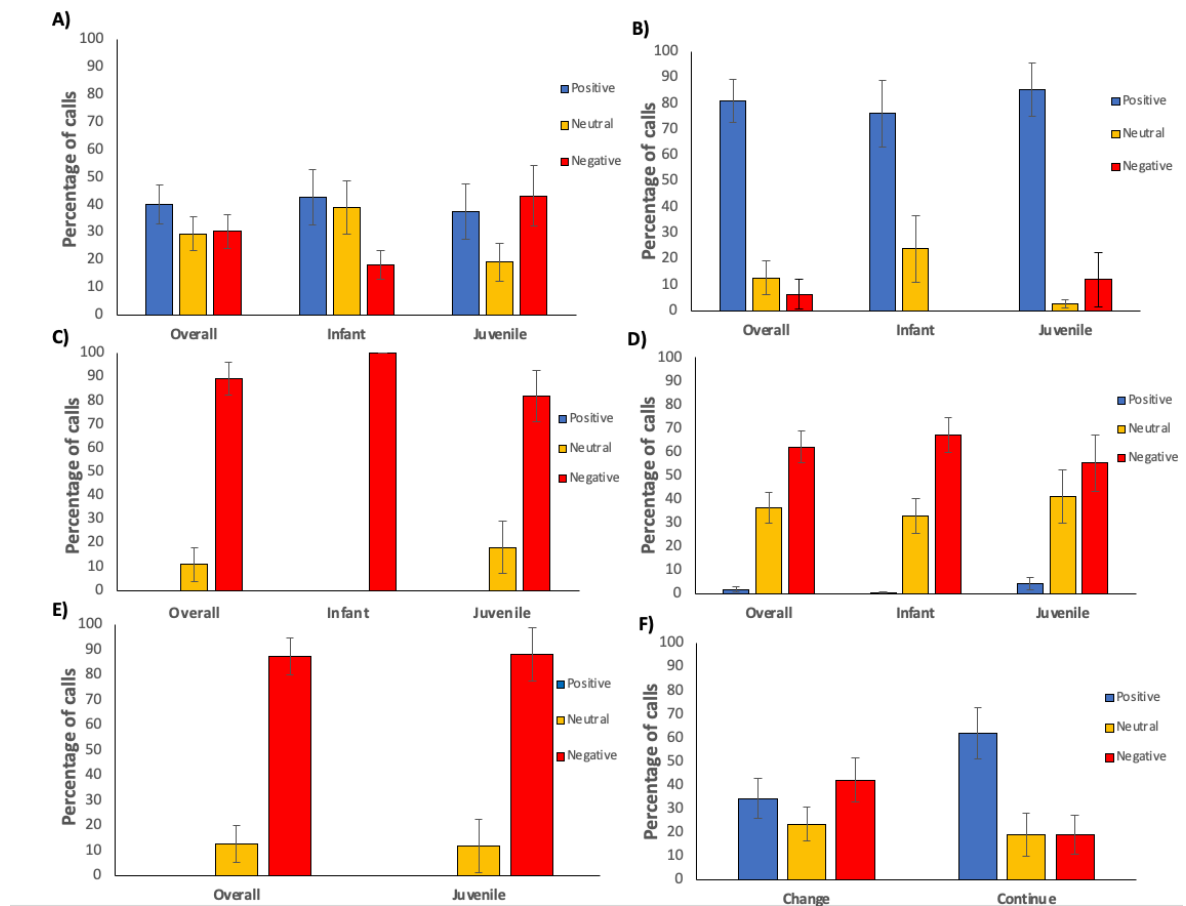


Figure 3.1 Bar graph depicting mean and standard errors of the percentage of grunts (A), laughs (B), screams (C), whimpers (D) and hoo calls (E) that occurred during positive, neutral and negative affective states. F depicts the mean and standard errors of the percentage of positive, neutral, and negative grunts that elicited behavioural change and continuation in a social partner.

Overall and for juveniles but not infants, there was a significant difference in the percentage of negative (Wilcoxon test, overall: $Z=-1.80$, $N=14$, $p=.017$; infants: $Z=-0.216$, $N=6$, $p=.138$; juveniles: $Z=-2.366$, $N=8$, $p=.007$) and neutral (Wilcoxon test, overall: $Z=-2.311$, $N=11$, $p=.020$; infants: $Z=-.733$, $N=6$, $p=.176$; juveniles: $Z=-2.02$, $N=5$, $p=.018$), grunts that elicited change vs continuation in a social partner, consistent with functional flexibility. However, there was no significant difference in the percentage of positive

grunts that occurred when social partners changed vs continued their behaviour

(Wilcoxon test, overall: $Z=-1.419$, $N=15$, $p=.155$; infants: $Z=-1.83$, $N=7$, $p=.133$; juveniles: $Z=-.070$, $N=8$, $p=.175$). Direct infant-juvenile comparisons of call functionality are shown in Appendix 2 supplementary results B.

There was a consistent significant difference in the affective valence of laughs (Friedman test, overall: $\chi^2(3)=18.67$, $N=17$, $p<.001$; infants: $\chi^2(3) = 8.85$, $N=8$, $p=.007$; juveniles: $\chi^2(3)=11.267$, $N=9$, $p<.001$), suggesting laughs are affectively biased (see figure 3.1B).

Post-hoc tests showed laughs were consistently significantly more likely to be positive vs negative (Wilcoxon test, overall: $Z=0.126$, $N=17$, $p<.001$; infants: $Z=1.312$, $N=8$, $p=.013$; juveniles: $Z=-.122$, $N=9$, $p=.007$). Overall and for juveniles but not infants, laughs were significantly more likely to be positive vs neutral (Wilcoxon test, overall: $Z=-1.029$, $N=17$, $p=.002$; infants: $Z=-.750$, $N=8$, $p=.100$; juveniles: $Z=-1.278$, $N=9$, $p=.010$). There was consistently no difference in neutral vs negative laughs (Wilcoxon test, overall: $Z=-0.23$, $N=17$, $p=.500$; infants: $Z=-.562$, $N=8$, $p=.260$; juveniles: $Z=0.56$, $N=9$, $p=.333$). Laughs did not show signs of functional flexibility (i.e. social partner responses were almost always associated with positive laughs and there was no difference in the percentage of laughs of each affective valence that were associated with change vs. continuation, see Appendix 2 supplementary results B)

There was a consistent significant difference in the affective valence of screams

(Friedman test, overall: $\chi^2(3)=22.73$, $N=15$, $p<.001$; infants: $\chi^2(3)=19.54$, $N=6$, $p<.001$; juveniles: $\chi^2(3)=11.64$, $N=9$, $p=.001$), suggesting screams are affectively biased (see figure 3.1C). Post-hoc tests showed that for infants, juveniles, and overall, screams were

significantly more likely to be associated with negative vs positive affective states (Wilcoxon test, overall: $Z=1.43$, $N=15$, $p<.001$; infants: $Z=1.50$, $N=6$, $p=0.28$; juveniles: $Z=1.389$, $N=9$, $p=.010$). Overall and for infants but not juveniles, screams were more likely to be negative vs neutral (Wilcoxon test, overall: $Z=1.16$, $N=15$, $p=.002$; infants: $Z=1.50$, $N=6$, $p=0.14$; juveniles: $Z=.944$, $N=9$, $p=.067$). Screams were not more likely to be associated with neutral vs positive affective states (Wilcoxon test, overall: $Z=.267$, $N=15$, $p=.333$; infants: $Z=.000$, $N=6$, $p=.333$; juveniles: $Z=.444$, $N=9$, $p=.333$). Screams did not show signs of functional flexibility (i.e. screams were only associated with behavioural change, which occurred mostly when screams were negative, see Appendix 2 supplementary results B).

There was a consistent significant difference in the affective valence of whimpers (Friedman test, overall: $\chi^2(3)=26.079$, $N=21$, $p<.001$; infants: $\chi^2(3)=18.565$, $N=12$, $p<.001$; juveniles: $\chi^2(3)=8.467$, $N=9$, $p=.007$), suggesting whimpers are affectively biased (see figure 3.1D). Post-hoc tests showed that overall, as well as for infants and juveniles, whimpers were more likely to be negative vs positive (Wilcoxon test, overall: $Z=1.048$, $N=21$, $p=.001$; infants: $Z=1.042$, $N=12$, $p=.016$; juveniles: $Z=1.056$, $N=9$, $p=.018$). Overall and for infants but not juveniles, whimpers were significantly more likely to be negative vs neutral (Wilcoxon test, overall: $Z=1.452$, $N=21$, $p=.001$; infants: $Z=1.708$, $N=12$, $p<.001$; juveniles: $Z=1.111$, $N=9$, $p=.075$). Whimpers were not more likely to be positive vs neutral for infants, juveniles, and overall (Wilcoxon test, overall: $Z=.405$, $N=21$, $p=.189$; infants: $Z=.667$, $N=12$, $p=.102$; juveniles: $Z=.056$, $N=9$, $p=.1.000$). Whimpers did not show signs of functional flexibility (i.e. whimpers were only associated with behavioural change, which occurred mostly when whimpers were negative, see Appendix 2 supplementary results B).

There was a significant difference in the affective valence of hoo calls overall and for juveniles but there was insufficient data to examine infants (N=4) (Friedman test, overall: $\chi^2(3)=12.950$, N=10, $p<.001$; juveniles: $\chi^2(3)=6.583$, N=6, $p=.018$), suggesting hoo calls are affectively biased (see figure 3.1E). Overall and for juveniles hoo calls were significantly more likely to be negative compared to neutral (Wilcoxon test, overall: $Z=-2.59$, N=10, $p=.009$; juveniles: $Z=1.838$, N=6, $p=.014$) or positive (Wilcoxon test, overall: $Z=-2.803$, N=10, $p=.005$; juveniles: $Z=2.121$, N=6, $p=.005$). However, hoo calls were not more likely to be neutral compared to positive (Wilcoxon test, overall: $Z=1.725$, N=10, $p=.077$; juveniles: $Z=1.00$, N=6, $p=.423$). Only 3 hoo calls were observed during social interactions. As such, hoo calls did not show signs of functional flexibility.

Discussion

The present study aimed to examine functional flexibility in chimpanzee vocal ontogeny. It was found that for both infants and juveniles, the vocal repertoire comprises a mixture of freely expressed and more affectively biased call types, much like human infants. Juveniles were, however, more stereotyped in how laughs were produced, but less stereotyped in how screams and whimpers were produced. Freely expressed calls (i.e. grunts) also showed stronger evidence for functional flexibility than more affectively biased call types, as social partner responses were to some extent dependent on the affective valence of the grunt, particularly among juveniles. Together, these results indicate that free expressivity and functional flexibility are present in the vocal behaviour of immature chimpanzees and suggest functional flexibility in particular may be a social competency that improves during ontogeny.

The hypothesis that grunts would be freely expressed whereas other call types would be more affectively biased was supported by the observation that grunts were regularly produced when the immature chimpanzees seemed to be in positive, neutral, or negative affective states, while laughs were predominantly linked to positive states and screams, whimpers, and hoo calls were predominantly linked to negative states. This finding replicates the findings of Dezechache and colleagues (2020) who also found grunts were freely expressed and significantly less affectively biased than whimpers, suggesting this is a reliable characteristic of immature chimpanzee vocal communication. By including additional call types that were not studied before but are believed to be homologous with human vocalisations (i.e. laughs, screams, whimpers, and hoo calls, see Davila Ross et al., 2009; Schwartz et al., 2020) our findings also strengthen the apparent parallel with human infant research (i.e. Oller et al., 2013) where it has been shown that laughs and cries are more affectively biased than protophones. This parallel suggests that free expressivity may be a precursor to language in phylogeny, as well as in ontogeny. Furthermore, we provide evidence that this communicative capacity persists until much later in ontogeny (at least until 10 years of age) than was shown by previous studies that focused on chimpanzees younger than 1 year of age (i.e. Dezechache et al., 2020). This in turn implies that functional flexibility is a pervasive social competency in chimpanzees as it is in humans, rather than being a peculiarity of early ontogeny.

Our data also supported the hypothesis that freely expressed vocal units are more functionally flexible (Oller, 2012) because grunts appeared to be freely expressed and social partner responses to grunts were dependent to some extent on the affective state

a grunt expressed. No such patterns were observed for affectively biased call types. This is the first study to show a non-human primate call type appears to be functionally flexible in the same sense that human infant protophones are functionally flexible (e.g. Oller et al., 2013) because previous primate studies did not examine social partner responses to freely expressed calls (e.g. Clay, Archbold & Zuberbühler, 2015; Dezecache et al., 2020). This helps to explain why freely expressed signals might be expected to be present in non-human primates, because if signals do not systematically effect receivers they are unlikely to be selected for (Maynard-Smith & Harper., 2003; Maynard-Smith & Harper, 1995). The observed pattern of social partner responses was arguably less pronounced than observed in human infants because behavioural change in a chimpanzee social partner was elicited by neutral and negative grunts *relative* to behavioural continuation, whereas for human infants positive protophones mostly elicited encouraging responses while negative protophones mostly elicited attempts to change the infants' state (e.g. Oller et al., 2013). This could reflect known species differences in co-operation (Moll & Tomasello, 2007). However, in human studies the social partner was always the subjects' caregiver, whereas social partners in chimpanzees included non-maternal group members who may have less prosocial motivations towards subjects due to weaker social bonds (Melis et al., 2006). Furthermore, studying caregiver responses to human infant vocalisations in the laboratory as was done in previous studies (e.g. Oller et al., 2013) may be subject to demand characteristics thereby exaggerating this apparent difference, as suggested by studies that have shown caregivers are more responsive to their infants when being observed in the lab compared to the home (Belsky, 1980). The evidence for functional flexibility in grunts was particularly strong in juveniles, suggesting that this may be a social competency that improves in ontogeny. Although, infants typically interact with their

mothers whereas juveniles interact with other group members (Plooij, 1984; Goodall, 1990), meaning the observed differences in call functionality could be attributable to a difference in social partner type than socio-communicative competency.

Laughs showed a strong positive affective bias and appeared to become increasingly stereotyped in the affective states they expressed in ontogeny. The overall positive bias in laughter was expected because human laughs are affectively biased (Oller et al., 2013) and seem to be homologous with chimpanzee laughs (Davila-Ross et al., 2009). Laughs also did not show signs of functional flexibility, which is consistent with the theory that affectively biased vocalisations cannot be functionally flexible (Oller, 2012). Our data did not permit us to enquire into why laughs might be more stereotyped in how they are produced and why this might increase during ontogeny. One explanation could be that laughter provides positive feedback to social partners to continue behaviours that might otherwise escalate into aggression during chimpanzee play (Matsusaka, 2004). Older chimpanzees are also known to be more likely to exhibit aggressive behaviour than younger individuals (Anestis, 2006). Consequently, laughs may be consistent in how they are produced to provide an unambiguous signal that mitigates the risks of aggression during play, and the importance of this increases in ontogeny due to increased risk of aggression. Interestingly, despite not being functionally flexible, laughs were the only call type other than grunts that elicited both behavioural change and continuation in a social partner. Since laughs are produced both spontaneously and in response to play actions (Davila-Ross et al., 2011), the observed patterns may reflect laughs being used to both initiate and maintain play, particularly among juveniles (see Appendix 2, supplementary results section B).

Screams showed a strong bias towards negative affective states. In contrast to laughs, screams were more consistent in how they were produced earlier in ontogeny. Similar to laughs, we expected to observe an affective bias in scream production because chimpanzee screams are believed to be homologous with those of humans, whose screams are also affectively biased (Schwartz et al., 2020). As predicted, there was also no evidence of functional flexibility in screams, providing further evidence supporting the theory that affectively biased vocalisations are less functionally flexible (Oller, 2012). Much like laughs, our data did not permit us to enquire into why screams were more consistent in how they were produced earlier in ontogeny. One explanation could be related to the fact that screams are often used in combination with other call types. For instance, screams are incorporated into pant-hoot calls (Crockford & Boesch, 2005), which are largely produced by older individuals (Fedurek et al., 2013), and are typically associated with states such as excitement (Goodall, 1990; Slocombe & Zuberbühler, 2010) rather than states that are more clearly negatively valenced such as fear (see Alarcao et al., 2018). The observed flexibility in scream production in older chimpanzees may therefore indicate increased control of scream production later in ontogeny which allows individuals to incorporate screams into different call sequences. Interestingly, an increased ability to control the production of affectively biased calls, such as screams, is also observed in human ontogeny (see Pisanski et al., 2016).

Much like laughs and screams, whimpers also showed a strong negative affective bias. This replicates the findings of Dezechache et al. (2020) who also found whimpers were biased towards negative affective states. However, our observations also build upon

Dezecache et al's (2020) study by showing that, similar to screams, whimpers were more consistent in their relationship to negative affective states earlier in ontogeny. Whimpers also showed no signs of functional flexibility, which was expected since whimpers were affectively biased in both our data and previous studies (Dezecache et al., 2020). The apparent differences in whimper production between infants and juveniles may be related to differences in the infant and juvenile ontogenetic niche. Whimpers are typically used during mother-offspring interactions and commonly elicit care-related maternal responses (Plooij, 1984). There are also high-levels of parent-offspring conflict during the juvenile weaning period whereby mothers attempt to reduce juvenile dependence by withholding care-giving behaviours such as breast-feeding or carrying (De Lathouwers & van Elsacker, 2006). Flexible whimper production in juveniles may allow more strategic use of whimpers to negotiate such parent-offspring conflict. Alternatively, juvenile whimpers may simply be less indicative of negative affect than infant whimpers, perhaps due to a lower dependency on maternal feeding (e.g. Pusey, 1983) which thereby reduces the severity of mothers withholding care-giving behaviours.

Hoo calls showed a strong bias towards negative affective states. However, unlike all other call types, hoo calls were produced almost exclusively by juveniles. Hoo calls have been previously reported to be incorporated in to whimper bouts (Goodall, 1990; Marler, 1976) as well as being produced in contexts of alarm, travel and rest (Crockford et al., 2018). The observation that hoo calls were only used in social interactions on 3 occasions, whereas whimpers were routinely used in social interactions, suggests that the observed hoo calls were most likely used to fulfil a broadcast function such as alarm rather than being used to regulate dyadic social interactions as grunts, laughs, screams, and

whimpers were used. Additionally, the observation that these calls were almost exclusively produced by juveniles is consistent with recent studies which have shown alarm hoo calls are produced more often in juvenile chimpanzees compared to infants (Dezecache et al., 2019). Our observations of hoo call production highlight an important point – free expressivity and functional flexibility is perhaps much more likely among calls that are used to navigate dyadic social interactions. This speculation may help to identify candidate species and call types in future studies of free expressivity and functional flexibility (i.e. calls used to regulate social interactions in species that show high levels of dyadic social behaviour).

The present findings on patterns of vocal production and function in immature chimpanzees may be limited by the system used for classifying affective states, because it is inevitably vulnerable to misclassification. Behavioural cues were central to affect classification in the present study. While many believe behavioural cues can communicate affective information (e.g. Dawkins, 2006), an alternative view is that such behaviours are ‘action-intention’ cues – cues that provide others with information about what an individual is likely to do next (Horstmann, 2003; Waller et al., 2017). On the action-intention view, behavioural cues are not necessarily indicative of any underlying affective state. This may be seen to question whether the present coding scheme classified affective state rather than another construct such as action-intentions, which would in turn question whether our data really shows free expressivity and functional flexibility. We make several arguments against this interpretation. Firstly, we do not see that these are mutually exclusive possibilities (i.e. a cue might indicate what an individual is likely to do because of the affective state associated with it). Secondly, a variety of sources of

empirical evidence do show that such behavioural cues are often associated with different affective states (see table 3.2). Therefore, while behavioural cues may not always be underpinned by affective states, they often are. As such, the present coding scheme is not considered to be a perfect system for classifying affective states, but we do argue it represents an improvement upon previous attempts that relied on a much smaller range of behaviours and did not include facial expressions (e.g. Clay, Archbold & Zuberbühler, 2015; Dezechache et al., 2020) which are among the most extensively studied and reliable affective cues (Cecilione et al., 2017). Importantly, chimpanzee facial expressions and bodily behaviours can be produced independently of vocalisations (Parr et al., 2007; Davila-Ross et al., 2011) and therefore provide independent evidence of affective state. Finally, it is important to be clear that the question at hand here is whether immature chimpanzees express different affective states, rather than whether those underlying affective states are truly occurring in a particular case. As such, we believe the present coding scheme was sufficient to provide reliable insights into affective expression of chimpanzee calls.

To conclude, free expressivity and functional flexibility appear to be present in the vocal communication of immature chimpanzees. Freely expressed calls appeared to be more functionally flexible, particularly among juveniles. A variety of differences were also observed in patterns of vocal production between infants and juveniles. These differences may be related to differences in the infant and juvenile ontogenetic niches, and also indicate that different call types have different ontogenetic trajectories. In human ontogeny, language is indeed built upon a pragmatic foundation of free expressivity and

functional flexibility. The present study suggests that language is built upon this foundation not only in ontogeny, but also in phylogeny.

4. Directedness and engagement in chimpanzee vocal ontogeny

Abstract

Directedness and engagement during pre-verbal vocal communication play a major role in language development. Did they also play a role in the ontogeny of evolutionarily earlier forms of communication? Understanding the evolutionary origins of language invites us to examine the role of these behaviours in the vocal ontogeny of chimpanzees, our closest living relatives. In the present study, we collected observational data on infant (N=13) and juvenile (N=15) chimpanzees from 0–10 years of age at Chimfunshi Wildlife Orphanage, Zambia. We examined the impact of age and vocalization type (grunts, whimpers, laughs and screams) on directed cues (gaze directedness and face directedness) and engagement (mutual gaze) during vocal communication. We also assessed the impact of directed cues and engagement on social interactions by coding the behaviour of social partners before, during, and after a vocalisation, and examining whether they contingently changed their behaviour in response to the vocalisation if it was directed or if engagement occurred. We found that face directed vocalisations showed a general increase during ontogeny and were predictive of behavioural responses in social partners, which was not the case for gaze directed vocalisations or vocalisations that occurred during mutual gaze wherein the effects of ontogeny were absent and dependent on call type respectively. We also found an increase in how consistent individuals were in the way their vocalisations were directed towards social partners in ontogeny, which was in turn associated with increased chances of eliciting behavioural responses from social partners. We conclude that similar to young humans, young

chimpanzees routinely exhibit directed behaviours and engagement during vocal communication. This social competency improves during ontogeny and benefits individuals by increasing the chances of eliciting behavioural responses from social partners. Directedness and engagement likely provide a foundation for language phylogenetically, as well as ontogenetically.

Introduction

Directedness of communicative acts (i.e. selection of a specific social target through cues such as gaze direction and facial orientation) and engagement during communication (i.e. mutual attention while communicating) are known to play an important role in language development (Lavelli & Fogel, 2005; Donnellan et al., 2020). While the reasons for this are not fully understood, it has been suggested that directedness and engagement during communication impact language development because these aspects of a communicative act facilitate the expression of intentions (Tomasello et al., 1997) and initiation of dyadic interactions (Bornstein et al., 1999). Human infants have been observed to flexibly direct their vocalisations towards specific individuals using gaze from 3 months of age and this is mostly linked with pre-speech sounds (i.e. protophones) rather than non-speech sounds such as laughter and cries (Oller, 2012). Since protophones are more flexibly produced than non-speech sounds and are pre-cursors to linguistic communication (Oller et al., 2013), this further supports the contention that directedness plays an important role in language development and also suggests flexibility in vocal production may be an essential pre-requisite for directed vocal communication (Vihman et al., 1985; Oller, 2012).

The role of directed vocal communication in early social interaction is demonstrated by studies using the still face paradigm, which have shown that, by 5 months of age, human infants seem to use directed vocalisations to capture the attention of an inattentive social partner (Delgado, Messinger, & Yalo, 2002; Goldstein, Schwade, & Bornstein, 2009).

However, studies that sampled a wider range of ages, from 4 – 10 months of age, found that infants did not direct their vocalisations towards others above chance level until 10 months of age, suggesting directed vocal behaviour is a social competence that improves during ontogeny (D’Odorico et al., 1997; D’Odorico & Cassibba, 1995). This may be related to maternal responsiveness, since infants whose mothers are more responsive to directed vocal behaviour at 8 months show significantly more directed vocal behaviour at 14 months (Gros-Louis et al., 2014). Among 11-month-old human infants, directed vocalisations are also more likely to elicit responses from caregivers than non-directed vocalisations, demonstrating that directed vocal behaviour is effective at engaging social partners (Donnellan et al., 2020). Moreover, infants who show more directed vocal behaviour at this age have been found to have a greater expressive vocabulary size at 2 years of age (Donnellan et al., 2020), demonstrating a direct relationship with language.

The engagement between infants and social partners established by directed communicative acts also impacts human communicative development. Engagement is important during early ontogeny because it provides a means of understanding others’ psychological states (Reddy, 2019), and while engagement may take many forms (i.e. tactile engagement, auditory engagement, etc), a particularly common form early in human ontogeny is visual engagement as indicated by mutual gaze (Reddy, 2010). As early as two months of age, vocalisations and facial expressions are routinely produced

during mutual gaze (Lavelli & Fogel, 2005), suggesting infants harness mutual engagement as an opportunity to communicate. This is further supported by the observation that infants aged 4 - 24 weeks produce vocalisations more frequently when participating in mutual gaze than when not (Hsu & Fogel, 2010) and by 4 months of age, infants show increased activity in the temporal and prefrontal cortex during mutual gaze (Grossmann et al., 2008). At this age, the vocalisations produced during mutual gaze mostly express positive affect (Colonnesi et al., 2012), which has been observed to encourage caregivers to continue their interactions with infants (Oller et al., 2013). Such forms of engagement have also been found to increase the diversity of call types infants produce (Franklin et al., 2014), suggesting an important role in vocal ontogeny. While directedness and engagement appear to play an important role in human communicative ontogeny, the role of such behaviours in the phylogeny of human language and communication is less known.

In the present study, we aimed to assess ‘directed’ vocal communication and visual engagement during vocal communication throughout chimpanzee ontogeny, using approaches (i.e. measuring subjects’ gaze direction and facial orientation towards social partners, as well as mutual gaze) that align with those used in the human infant literature (e.g. Franklin et al., 2014; Donellan et al., 2020). Indeed, there is good reason to expect such visual cues to play a role in chimpanzee communication. Aspects of primate eye-morphology that are known to be important during communication, such as width of the iris, strongly correlate with social complexity among primates (Kobayashi & Hashiya, 2011). Since chimpanzee social structure has been shown to outstrip that of many other primate species regarding complexity (Aureli & Schino, 2018), this suggests that gaze-

related behaviours likely play an important role in chimpanzee communication. Studies of mother-infant interactions in chimpanzees have shown that mother-infant dyads routinely participate in mutual gaze (Bard, 1994) and there are group level differences in this pattern which may relate to the environment of different populations (Bard et al., 2005). To our knowledge, however, there are no studies of the relationship between mutual gaze and vocal communication in chimpanzees.

Directedness has played an important role in research on ape gestural communication (see Krause et al., 2018; Tomaello & Call, 2019). However, a comparably small number of studies have examined whether non-human primate vocalisations are ‘directed’ to specific individuals. Seyfarth & Cheney (2018) reported that baboons direct calls to specific individuals. For example, according to these authors, baboon grunts are usually directed towards lower ranking individuals, and behavioural cues such as gaze can be used to ascertain who the target of a call is, indicating directedness may also play an important role in primate vocal communication. The evidence on directed vocal communication among great apes is more indirect inasmuch as behavioural cues of directedness have seldom been studied. Instead, playback experiments are typically used. For example, playback experiments with chimpanzees and bonobos show that individuals selectively produce vocalisations in response to hearing vocalisations from specific individuals (i.e. those who are high ranking: Schel et al., 2013; Genty et al., 2013) and selectively vocalise in response to conspecifics who are unaware of the presence of a predator (Crockford et al., 2012, 2017). These studies hint at the possibility of directed vocal communication in great apes. However, since these studies differ markedly in their

methodology from studies of human infants, the implications of such studies for our understanding of the evolutionary origins of language remain unclear.

With regard to ontogeny, Laporte & Zuberbühler (2011) observed that similar to human infants, grunts began to be directed towards others at approximately 2 months of age. Later in ontogeny (i.e. during the juvenile and sub-adult periods), however, grunts were directed towards others less often, but increased in the specificity of who grunts were directed towards – primarily dominant males. Unlike the human literature though, Laporte & Zuberbühler (2011) considered vocalisations to be directed if produced when a conspecific approached but did not examine communicative behaviours such as face directedness. Interestingly, the authors also reported never observing vocal communication during mutual gaze. Dezechache, Crockford, & Zuberbühler (2019) studied the role of gaze in the ontogeny of chimpanzee alarm calls. The authors found alarm calling increased with age, but gaze alternations between conspecifics and a predatory stimulus during alarm calls were common at all ages, suggesting alarm call ontogeny may be influenced by early gaze alternations. Notably, both studies focused on a single call type, whereas a key prediction in the human infant literature is that directedness is primarily related to more flexibly produced call types (Oller, 2012), highlighting the importance of comparisons between call types. No study of non-human primate vocal production has yet examined the ontogeny of directed behaviours during communication, such as gaze directedness, and more notably face directedness and mutual gaze, which play a particularly important role in human communicative acts (Gomez, 1996; Moore, 2016). Furthermore, while in humans, directed communicative behaviours are more likely to elicit responses from social partners (Donnellan et al., 2020), no studies of

chimpanzees have examined the impact of directed communicative behaviours on social interactions.

To fill this gap in the literature, we assessed directed vocal communication and engagement in semi-wild infant and juvenile chimpanzees. We aimed to assess both how often directedness and engagement occur during vocal communication in chimpanzee ontogeny, and how flexibly vocalisations are directed, which has been reported to be a characteristic of pre-vebal vocal communication in human infants (Oller, 2012). We also assessed the impact of directed communication and engagement during communication on social interactions by testing whether calls that are directed or occur during engagement are more likely to elicit behavioural responses from social partners. Since directed communication is related to flexibility in vocal production among human infants (Oller, 2012), and grunts are more flexibly produced than other call types in young chimpanzees (Dezecache et al., 2020; Taylor et al., under review), we predicted directedness and engagement, and any developmental changes associated with these aspects of communication, would primarily be related to grunts. Achieving these aims will fill a crucial gap in our understanding of the evolutionary origins of language by elucidating whether directedness and engagement play a role in chimpanzee vocal ontogeny as is observed in humans.

Method

Subjects and study site

Subjects were infant (N=13) and juvenile (N=15) chimpanzees housed at Chimfunshi Wildlife Orphanage (CWO), Zambia. Infant ages ranged between 0 – 4 years of age ($M=1.13$, $\pm SD=1.14$). Juvenile ages ranged from 4 – 10 years of age ($M=7.07$, $\pm SD=3.82$). All subjects were raised by their mothers during infancy, and all infant subjects still lived with their biological mothers. Three juveniles (one 5 year old male, one 10 year old male, and one 8 year old female) did not live with their mothers due to fatalities that occurred in years prior to the present study. While the majority of chimpanzees at Chimfunshi are rescued from adverse circumstances, such as the pet trade, all subjects in the present study were born in captivity. However, the majority (78%) of their mothers were originally wild. Chimpanzees housed at the CWO belong to a mixture of chimpanzee sub-species, including *Pan troglodytes troglodytes* and *Pan troglodytes schweinfurthii*. Each subject belongs to one of four mixed-sex colonies that comprise between 10 – 52 members. For a further breakdown of the study population see table 4.1 below.

Table 4.1 Overview of subject characteristics including developmental stage, age, sex, and group membership.

Developmental stage	Age (Years)	Sex	Group
Infant	0	Males (3); Females (2)	1, 2
	1	Males (0); Females (1)	2
	2	Males (3); Females (2)	1, 2
	3	Males (0); Females (2)	2
Juvenile	4	Males (1); Females (2)	2, 3

5	Males (1); Females (1)	2, 4
5	Males (1); Females (0)	4
6	Males (3); Females (1)	1, 2
7	Males (2); Females (0)	2, 4
8	Males (0); Females (1)	3
10	Males (2); Females (0)	3, 4

Each subject was housed in one of four outdoor miombo forest enclosures. Miombo woodland is the habitat of many wild chimpanzee populations (Schoeninger, Moore, & Sept, 1999). Enclosure sizes were 190 (group 1), 160 (group 2), 47 (group 3), and 62 (group 4) acres. Here, chimpanzees are fed once daily at approximately 12pm, and some chimpanzees are fed indoors. Beyond this, chimpanzees of CWO forage on naturally fruiting trees. The enclosures and group sizes are large enough that fission fusion social dynamics take place as well as regular territory patrols. At night, all chimpanzees sleep outside and many show nesting behaviours.

Data collection

Video and audio recordings were collected between 7am and 6pm from June 2018 to October 2018 (excluding 12pm-1pm when daily feeding occurred) using a Sony CX405 Handycam with a Sennheiser ME66 directional microphone attached. Recordings were collected only when the subjects were outdoors and the recordist was within 2–10 meters of the subject. The main approach in collecting recordings was to use a 5-minute focal sampling method that allowed us to have equal representation of the sample in this study. These focal recordings for each subject were collected in a new order each day

determined by using a random number generator. However, due to the large size of the enclosures and the dense forest inside them, subjects were often not visible. For efficiency, we therefore decided to wait for 5 minutes to observe a subject. If the subject was not visible, we then recorded the visible chimpanzee who was next highest on the focal recording list. It was attempted to obtain two such focal recordings on a subject in a single day – one in the morning (before 12pm) and one in the afternoon (after 1pm). 650 5-minute focal recordings were collected overall. In addition, there were 44 recordings that were incomplete as the subjects left the view of the camera for more than 30 seconds and could therefore no longer be seen and identified as the potential caller. Furthermore, when there were no visible subjects where 5-minute focal recordings could be taken for that day, the infants were recorded ad libitum. 79 ad libitum recordings were taken. These additional recordings were also included in the analysis of this study in order to increase the overall number of calls. All data collection was ethically approved and permitted by the University of Portsmouth Animal Welfare and Ethical Review Body (AWERB) and Chimfunshi Research Advisory Board (CRAB) (see appendices 4 & 5).

Overall, between 15 and 51 focal recordings were collected per subject, meaning focal observation time ranged between 1.24 and 4.25 hours of recording per subject (Infants: $M=2.72 \pm SD=0.96$; Juveniles: $M=3.43 \pm SD=0.04$). The total duration of ad libitum recordings per subject ranged between 0.03 hours and 3.52 hours (Infants: $M=0.81 \pm SD=0.79$; Juveniles: $M=0.18 \pm SD=0.15$). The total duration of incomplete focal recordings ranged between 0.20 hours and 1.27 hours per subject (Infants: $M=0.70 \pm SD=0.33$; Juveniles: $M=0.55 \pm SD=0.28$). Overall, total observation time ranged between 1.73 hours and 5.45 hours per subject (Infants: $M=4.19 \pm SD=1.31$; Juveniles: $M=4.16 \pm SD=0.53$).

Identifying calls

The coded unit of vocal behaviour was the call type, which is a broad category of calls (i.e. grunts) which contains distinct variants (i.e. food grunt, pant grunt etc). Call types were chosen because there is wide agreement regarding the call types produced by immature chimpanzees (Goodall, 1990; Plooi, 1984; Slocombe & Zuberbühler, 2010) but whether immature chimpanzees exhibit distinct subtypes (i.e. food grunts, pant grunts, etc) is currently unclear due to a lack of systematic study. Moreover, in a recent study of infant and juvenile vocal behaviour (Taylor et al., under review), we demonstrated the validity of these call types by using unsupervised machine learning techniques to show that based on acoustic information alone and without call labels, a clustering algorithm grouped the majority of each call type into distinct call clusters, suggesting call types can be reliably distinguished from one another in infants and juveniles.

Calls could be comprised of a single call element, or a series of call elements otherwise known as a call 'bout'. 768 calls were identified in total. The call types included grunts (N=382), whimpers (N=147), laughter (N=139), screams (N=41), hoo calls (N=41), barks (N=8), squeaks (N=6), and pant hoots (N=4). There was sufficient data only for analyses of grunts, whimpers, laughs, and screams. Hoo calls were not included because directedness could only be coded for just 21 calls and only 5 of these were observed among infants, meaning there was insufficient data to test for an effect of ontogeny. Pant-hoots were also not included because they are a combination of call types rather than a single call type (Fedurek et al., 2013). Calls were identified based on auditory cues followed by systematic visual inspection of spectrograms according to the definitions used in Taylor et

al (under review) (see chapter 2). An inter-rater reliability test was performed on 20% of the total identified calls, and Cohen's Kappa revealed a good (See Cohen, 1960) level of reliability ($K=0.75$).

Directedness and engagement coding

Directedness and engagement were coded while a subject vocalised. A call was coded as directed or during engagement if at any time throughout its duration it met criteria for directedness or engagement. Thus, directedness and engagement did not need to occur throughout the entire call. While this may mean directedness and engagement are more likely to be observed for longer call bouts, we found there is no significant difference in total observed duration between calls (Friedman test: $\chi^2(4)=2.314$, $N=7$, $p=.509$).

Importantly, there was also no correlation between age and total duration observed vocalising (Spearman's Rho: $r_s(11)=.490$, $N=28$, $p=.149$). As such, call duration does not confound analyses of the relationship between age or call type and directedness or engagement.

Directedness

Two different types of directedness were studied. 1) Subjects' gaze is directed towards a conspecific while the subjects' body is orientated towards the conspecific (hereafter 'gaze directed'). 2) Subjects' face is orientated towards the face of a conspecific while the subjects' body is orientated towards the conspecific (hereafter 'face directed'). To be considered face directed, the subjects' face must have been judged to be within $< 45^\circ$ of the social partners face. Definitions 1 and 2 are important distinctions firstly because chimpanzees are known to be sensitive to both gaze and face orientation of others during

communication (e.g. Kaminski, Call & Tomasello, 2004). Secondly, following definition 1, subjects' may not necessarily be within the conspecifics line of sight during gaze directed vocalisations, meaning the conspecifics may not realise they are the target of a directed call, making definition 2 important to include. Finally, definition 2 is also important because the development of face-to-face vocal communication is a key milestone in human vocal development (Oller, 2012). If the face of the subject was not visible (i.e. the subject was out of view or facing the opposite direction to the recordist) directedness was not coded. All coding of directedness was binary – either the call was gaze directed or not, face directed or not. A high level of intra-rater reliability was achieved based on 20% of the full dataset (Gaze directed, $K=0.81$; Face directed, $K=0.84$)

Mutual gaze

Directed vocal behaviour is an essential pre-requisite for engaging others during communication when engagement is defined as mutual gaze (Bard et al., 2005) during communication. In this study, mutual gaze was considered to occur during communication if the subject's face was orientated towards the face of a social partner and the face of the social partner was orientated towards the face of the subject at the same time while the subject was vocalising. To be considered as face orientated for both subjects and social partners, both individuals faces must have been judged to be $< 45^\circ$ of one another. If the face of the subject or social partner was not visible (i.e. the subject or social partner was out of view, or facing the opposite direction to the recordist), mutual gaze was not coded. Mutual gaze coding was binary – either mutual gaze occurred during communication or not. A high level of intra-rater reliability was achieved based on 20% of the full dataset ($K=0.90$).

Social partner behaviour and responses

The social partner referred to the individual(s) that the subject was interacting with while vocalising. The behaviour of the social partner was coded 4 seconds before the vocalisation, during the vocalisation, and 4 seconds after the vocalisation. This time frame was chosen because studies of chimpanzee infants that examined mothers behavioural responses to their infants vocalisations used a similar time frame (e.g. Dezechache et al., 2020). All observed behaviours were coded using the ethogram developed by Plooij (1984). See table 4.2 below for behaviour categories and the behaviours that comprised each category. Behavioural responses in a social partner were defined as when a social partner either stopped a category of behaviour or started showing a new category of behaviour during or after the subjects’ vocalisation was produced. Notably, behavioural categories were not mutually exclusive (i.e. a social partner may show both protecting and comforting behaviours during a vocalisation), meaning if only one category stopped or was introduced while another was consistent throughout, the social partner would still be considered to have changed their behaviour. All coded behaviours were directed towards the subject. The social partners’ behaviour was intra-rater reliability tested using Cohen’s Kappa, which showed an excellent degree of reliability was achieved based on 20% of the overall dataset (K=0.88).

Table 4.2 Categories of social partner behaviour towards subjects and behaviours that belong to each category. See Plooij (1984) for definitions of behaviours.

Category	Behaviours toward subject
----------	---------------------------

Feeding offspring	Food sharing, breast feeding
Protecting	Defending subject, gathering subject
Comforting	Cradling, patting, embracing, bite-kissing, holding hand, extending hand
Playing	Play wrestling, tickling
Grooming	Grooming, inspecting
Approach	Approaching subject
Travel	Lowering back for subject to climb on, following subject, carrying subject
Avoidance	Avoiding subject, leaving subject, or breaking contact with subject outside of play
Preventing breastfeeding	Covering nipple (mothers only)
Threat	Arm raising gesture, biting at (outside of play), dominance displays
Taking	Taking or pulling objects in subjects possession
Causing discomfort	Pulling, slapping, hitting, dragging, pulling hair, pushing subject (outside of play)

Statistical analysis

Generalized linear mixed models with a binomial error structure, a log link function, and a Nelder-Mead optimizer were used to examine the impact of age (in years) and call type (grunts, whimpers, laughs, and screams) on directedness and engagement. Three models

were built with identical fixed effects and random effects structure, but different dependent variables that reflect different aspects of directedness and engagement. In one model the dependent variable was gaze directedness, in another the dependent variable was face directedness, and in the third, the dependent variable was mutual gaze (see 'directedness coding' above for definitions). In all models the fixed effects were age and call type. Since there were individual differences in the number of observations, ID was used as a random effect. Additionally, previous studies have shown different colonies of the study population significantly differ in aspects of their social behaviour such as social tolerance (see van Leeuwen et al., 2018) which is measured using social proximity and therefore might limit an individuals' ability to exhibit directed communicative behaviours. Consequently, we included colony as a random effect. Since individuals are nested within different enclosures, a nested random effects structure was used. Initially, sex was also included as a random effect as previous studies show young male and female chimpanzees are socialised differently (see Lonsdorf et al., 2014). However, when sex was included as a random effect, models were unable to converge because there was zero variance in parameter estimates for sex. Consequently, sex was not included as a random effect in any model. All models successfully converged and took the following form: Directedness as dependent variable, the interaction of Age in Years and Call Type as fixed effect, and the nested random effect of ID nested in Enclosure Number. This model was compared to both a reduced model without the main interaction and a null (intercept and random factor only) model. We used likelihood ratio tests to compare models and calculated conditional R squared values to measure the difference in explained variance between models. Wald Chi-Square tests were used to test whether each parameter in the final model was significant.

To assess the impact of directedness on social interactions, we used three generalized linear mixed models with a binomial error structure, a log link function, and a Nelder-Mead optimizer. Each model had the same dependent variable – whether or not a social partner changed their behaviour in response to a vocalisation. Each model had a single fixed effect which differed between each model. The fixed effects were either gaze directedness, face directedness, or mutual gaze. Two random effects were used. One random effect was ID, due to an uneven number of observations per individual. The second was call type due to differences in responsiveness to different call types, for instance, mothers show high levels of responsiveness to infant whimpers (see Dezechache et al., 2020). Social partner ID was originally nested within enclosure as per the models discussed above. However, this was removed from all models built to predict social partner responsiveness due to zero variance in parameter estimates for enclosure number in these models, which prevented model convergence. All models converged and took the following form: Social partners' response as dependent variable, Directedness as fixed effect, and ID and Call Type as random effects. These were compared to a null (intercept plus random effects only) model. Likelihood ratio tests were employed to compare models and conditional R squared was calculated to measure the difference in explained variance between models. Wald Chi-Square tests were used to test whether each parameter in the final model was significant.

All generalised linear mixed models were run in R V3.5.2 (R Core Team, 2016) using the lme4 package V1.1.23 (Bates et al., 2007) to fit the models. The emmeans package V1.5.1 (Lenth et al., 2018) was used to perform post-hoc comparisons with Tukey comparisons in

the best fitting model. The MuMIn package V1.43.17 (Barton & Barton, 2015) was used to calculate conditional R squared values. All models were also subject to checks to ensure the models were a good fit to the data and did not violate assumptions of the models (see Appendix 3 supplementary methods).

To measure flexibility in how vocalisations were directed towards others, KR20 values were calculated for each individual. The KR20 formula provides a measurement of consistency for polychotomous variables analogous to a Cronbach's Alpha (see Kuder & Richardson, 1937). For each vocalisation, it was coded whether or not subjects were 1) gazing towards a social partner, 2) their torso was orientated towards the torso of a social partner, 3) whether they gazed at the social partner while their body was orientated towards the social partner, and 4) whether their face was directed towards the face of the social partner. KR20 values range between 0 and 1. A low KR20 value indicates subjects are inconsistent in how they direct their vocalisations towards social partners and are therefore flexible in how they direct their vocalisations. By contrast, a high KR20 value indicates subjects are consistent in how they direct their vocalisations towards social partners, and therefore suggests less flexibility. Notably, more consistency in directedness does not necessarily correspond to more directedness, because an individual could use only a single cue of directedness but be highly consistent in their use of that cue.

Censored regression was used with the censReg V0.5-32 package in R (Henningsen, 2020) to examine the influence of age and call type on consistency in how vocalisations were directed towards social partners. This method was chosen because the distribution of

KR20 was heavily left skewed and contained 0's. Censored regression was also used to examine whether consistency in how vocalisations are directed towards social partners impacts call functionality. For examining the ontogeny of consistency in directedness a full model containing an interaction between Age in Years and Call Type as the fixed effect was compared to a reduced model without the interaction and a null model (intercept only) using likelihood ratio tests. Wald Chi-Square tests were used to test whether each parameter in the final model was significant. To examine whether consistency influenced the probability of eliciting behavioural change in social partners a simple model with KR20 as the sole predictor of social partner responses was compared to a null model using a likelihood ratio test.

Results

Ontogeny of directedness & engagement during vocal communication

A full model wherein gaze directedness was predicted by an interaction between age and call type was significantly different from a null model (LRT: $\chi^2_7 = 27.136$, $p < .001$, $R^2_c = .180$). However, the full model was not significantly different from a reduced model including age and call type as fixed effects with no interaction between them (LRT: $\chi^2_3 = 5.433$, $p = .143$). As such, the reduced model was preferred. In this reduced model, there was no significant effect of age on gaze directedness during vocal communication. However, there was a significant effect of call type (Wald Chi-Square test: $\chi^2_3 = 5.433$, $p < .001$), with gaze directedness during communication being significantly more likely for screams ($\beta = 1.508$, $SE = 0.451$, $Z = 3.343$, $p = .004$), and whimpers ($\beta = 0.821$, $SE = 0.244$, $Z = 3.363$, $p = .004$), compared to grunts (Figure 4.1A). All model values for the preferred model are shown in Appendix 3 supplementary results A.

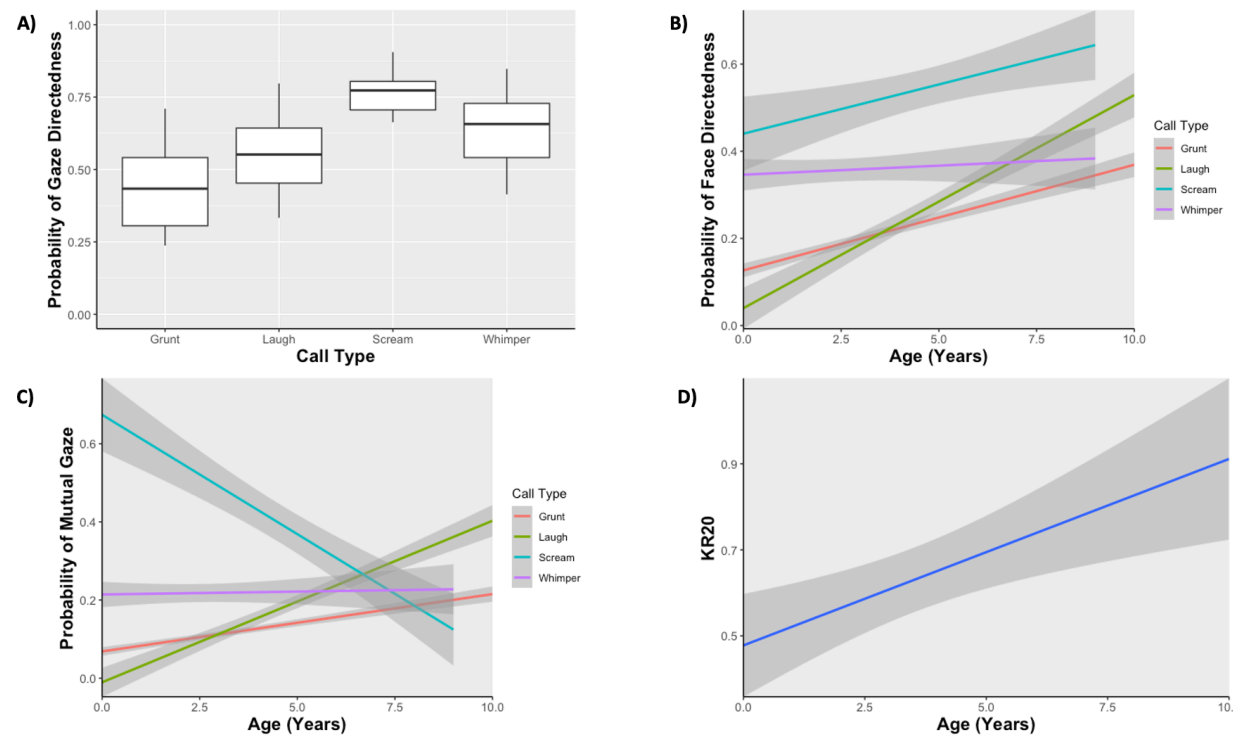


Figure 4.1 Main effects of models using age and call type as predictors of gaze directedness (A), face directedness (B), mutual gaze (C), and consistency in directedness (D) during vocal communication. Grey areas indicate 95% confidence intervals.

A full model wherein face directedness was predicted by an interaction between age and call type was significantly different from a null model (LRT: $\chi^2_7 = 25.919$, $p < .001$, $R^2_c = .230$). The full model was also significantly different from a reduced model wherein face directedness was predicted by age and call type with no interaction (LRT: $\chi^2_3 = 8.877$, $p = .030$). As such, the full model was preferred. In this model, there was a significant main effect of age (Wald Chi-Square test: $\chi^2_1 = 4.082$, $p = .043$), with individuals being increasingly likely to direct vocalisations towards the face of a social partner with increasing age ($\beta = 0.148$, $SE = 0.073$, $Z = 2.020$, $p = .043$). There was also a significant main effect of call type (Wald Chi-Square test: $\chi^2_3 = 17.708$, $p < .001$), with vocalisations

being significantly more likely to be directed towards the face of a social partner if they were whimpering ($\beta = 0.797$, $SE = 0.274$, $Z = 2.915$, $p = .018$) or screaming ($\beta = 1.166$, $SE = 0.453$, $Z = 2.577$, $p = .049$) compared to grunting. However, a significant interaction was also observed between age and call type (Wald Chi-Square test: $\chi^2_3 = 8.445$, $p = .037$) with laughs becoming increasingly likely to be directed towards the face of a social partner with age relative to whimpers ($\beta = 0.381$, $SE = 0.133$, $Z = 2.861$, $p = .022$) (Figure 4.1B). All model values for the preferred model are shown in Appendix 3 supplementary results B.

A full model that predicted mutual gaze during vocalisation using an interaction between age and call type was significantly different to a null model (LRT: $\chi^2_7 = 24.326$, $p < .001$, $R^2_c = .276$). The full model was also significantly different from a reduced model that included age and call type as fixed effects with no interaction between them (LRT: $\chi^2_3 = 12.84$, $p < .001$). The full model was therefore preferred. In the full model, there was no significant main effect of age. However, there was a significant main effect of call type (Wald Chi-Square test: $\chi^2_3 = 18.823$, $p < .001$), with mutual gaze being significantly more likely to occur if the vocalisation was a scream compared to a laugh ($\beta = 1.779$, $SE = 0.593$, $Z = 3.002$, $p = .014$), or a grunt ($\beta = 1.710$, $SE = 0.513$, $Z = 3.335$, $p = .004$). There was, however, also a significant interaction between age and call type (Wald Chi-Square test: $\chi^2_3 = 11.729$, $p = .008$) with screams being significantly less likely to occur during mutual gaze as age increased relative to laughs ($\beta = 0.629$, $SE = 0.200$, $Z = 3.145$, $p = .009$) (Figure 4.1C). All model values for the preferred model are shown in Appendix 3 supplementary results C.

To test for ontogenetic changes in how flexibly vocalisations were directed towards social partners, a censored regression model was built with an interaction between age and call types as predictors of KR20 values, which measure consistency in how vocalisations were directed towards social partners. This model was not significantly different from a null model (LRT: $\chi^2_3 = 10.823$, $p = .055$). However, a reduced model with no interaction was significantly different from a null model (LRT: $\chi^2_3 = 10.192$, $p = .017$). This reduced model was therefore preferred and showed a significant main effect of age (Wald Chi-Square test: $\chi^2_1 = 10.461$, $p = .001$) but not call type, suggesting that as individuals become older, they become more consistent in how they direct their vocalisations towards social partners ($\beta = 0.044$, $SE = 0.013$, $t = 3.263$, $p = .001$). (Figure 4.1D). All model values for the preferred model are shown in Appendix 3 supplementary results D.

Function of directedness & engagement during vocal communication

A model using gaze directedness to predict whether or not a social partner would change their behaviour in response to a vocalisation was not significantly different from a null model (LRT: $\chi^2_1 = 0.405$, $p = .524$, $R^2_c = .363$), suggesting gaze directedness does not modify the chances that a vocalisation will elicit a behavioural response from a social partner. However, a model using face directedness to predict whether or not a social partner would change their behaviour in response to a vocalisation was significantly different from a null model (LRT: $\chi^2_1 = 4.209$, $p = .040$, $R^2_c = .335$) (Figure 4.2A), suggesting that if a vocalisation is directed towards the face of a social partner, it is more likely to elicit behavioural change ($\beta = 0.553$, $SE = 0.267$, $Z = 2.068$, $p = .038$). Similar to gaze directedness, a model using mutual gaze to predict whether or not a social partner would

change their behaviour in response to a vocalisation was not significantly different from a null model (LRT: $\chi^2_1 = 1.182$, $p = .276$, $R^2_c = .343$). A censored regression showed subjects' who directed their vocalisations towards social partners in more consistent ways as measured by KR20 values were significantly more likely to elicit behavioural responses in a social partner compared to a null model (LRT: $\chi^2_1 = 7.640$, $p = .005$) (Figure 4.2B), suggesting effectively eliciting behavioural changes in social partners is partially dependent on adopting a consistent social style ($\beta = 0.300$, $SE = 0.108$, $t = 2.764$, $p = .005$).

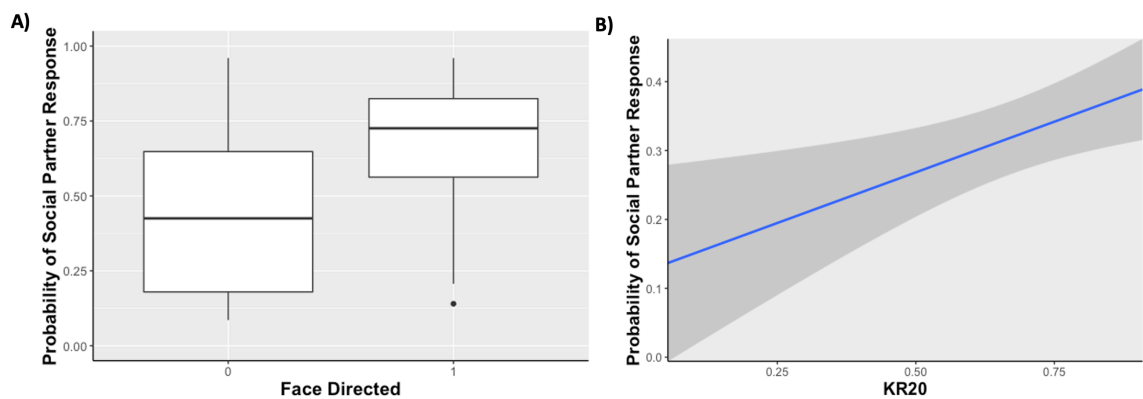


Figure 4.2 Main effects of models that successfully predicted social partner responses with face directedness (0 = not face directed, 1 = face directed) (A) and consistency (i.e. KR20 values) (B) as predictors. Grey areas indicate 95% confidence intervals.

Discussion

Directedness and engagement during vocal communication are fundamental for the emergence of language. This study aimed to assess the ontogeny of directed vocal communication and engagement during vocal communication in chimpanzees, and evaluate its impact on social interactions. We found that young chimpanzees routinely show signs of directedness and engagement during vocal communication, but only face

directedness showed a general increase during ontogeny. Interestingly, face directedness was also the only cue that increased the chances of a call eliciting a behavioural response from a social partner. All other ontogenetic changes in directedness and engagement during vocal communication depended on the call type. We also aimed to assess the ontogeny of flexibility in the use of directed cues during vocal communication, finding that chimpanzees become increasingly consistent in how they direct their vocalisations towards social partners throughout ontogeny. Those who used directed vocalizations in a more consistent manner were also more likely to elicit behavioural changes in social partners during vocal communication.

Much like human infants indeed, chimpanzees routinely directed their vocalisations and visually engaged with others during vocal communication from a young age. Fifty-six percent of vocalisations were directed towards a conspecific using gaze, 34% of vocalisations were directed towards the face of a conspecific, and mutual gaze occurred during 17% of vocalisations. Gaze directedness, face directedness, and mutual gaze during grunts and whimpers were all observed within the first month of life, with directedness and mutual gaze for all other call types being observable within the first 6 months. This observation contrasts somewhat with Laporte and Zuberbühler's (2011) study of wild chimpanzee grunts during the first 15 years of life, wherein it was reported that vocal communication during mutual gaze did not occur, but 'directed' vocalisations were observed from 2 months of age onwards. Consistent with our findings, previous (non-vocalisation) research on chimpanzee infants within the first 3 months of life has shown that chimpanzees show mutual gaze within the first month, but population differences emerged later from 2 months onwards, which appeared to be related to

differences in the amount of tactile social interaction (Bard et al., 2005). This suggests inconsistencies across studies may represent legitimate differences (rather than non-replications). This also raises the important point that engagement across human cultures is not universally visual and often tactile instead (e.g. Negayama et al., 2015), indicating that while engagement may be important for language development, *visual* engagement per se may not be essential. Further primate studies could also explore the impact of such forms of engagement on communicative development.

Face directed vocal behaviour did increase generally (i.e. independent of call type) during ontogeny and also was followed by an increased probability of eliciting behavioural responses in a social partner. This pattern was not observed for gaze directedness. Nonetheless, this provides evidence that the use of directed cues is a social competency that improves in ontogeny and is important for effective communication, as is observed in human vocal ontogeny (see Donnellan et al., 2020). Gaze directedness does not require the subject to be within the line of sight of a recipient (i.e. the subject could be behind a recipient and gaze at them while vocalising), whereas face directed communication is much more likely to be within the line of sight of a recipient. This may help to explain why calls with gaze directedness were not more likely to elicit behavioural responses – often recipients may not know they are the target. However, several important questions remain. Firstly, it is unclear why face directedness during vocal production elicits behavioural responses in social partners. One explanation is that face directedness may provide additional information for the interpretation of vocalisations through concurrent facial behaviours, which can indicate an individuals' affective state in chimpanzees (Parr et al., 2007), thereby enhancing a social partners' ability to understand the subjects'

communicative act and in turn respond to it. Alternatively, since face directed gaze increases arousal levels in humans (Hietanen et al., 2008; Pönkänen et al., 2011), the increased chances of eliciting behavioural change in a social partner could be arousal-mediated. Secondly, it is not clear why face directedness during vocal production increased throughout chimpanzee ontogeny. In human infants, those who have mothers that are more responsive to vocal communication show more directed vocal behaviour later in ontogeny (Gros-Louis et al., 2014), which in turn influences language development (Donnellan et al., 2020). However, we were not able to assess these possibilities in the present study because they require a longitudinal design (see D’Odorico et al., 1997). While time-consuming, such datasets can be certainly acquired with more systematic monitoring of mother-infant dyads.

Face directedness, but not gaze directedness, showed an interaction between age and call type, with laughs becoming significantly more face directed with age relative to whimpers. Thus, while the use of directed cues for whimpers was present from early in ontogeny, the use of directed cues was a competency that emerged in ontogeny for laughs in particular. This contrasts our prediction that ontogenetic changes in directedness would be primarily related to grunts, and in turn implies that early flexibility in vocal production, which is characteristic of grunts (Dezecache et al., 2020; Taylor et al., in prep), is not essential for the development of directed communication as has been suggested (see Oller, 2012). Studies of laughter during chimpanzee play show that laughter provides positive feedback to social partners to continue behaviours that might otherwise escalate into aggression (Matsusaka, 2004). Older chimpanzees are also known to be more likely to exhibit aggressive behaviour than younger individuals (Anestis, 2006).

Since directed forms of communication can reduce ambiguity in communication (Scott-Phillips, 2015), the use of directed cues during laughter may reduce ambiguity during playful interactions, which becomes increasingly important with age.

Both face and gaze directedness did show a similar effect of call type, mostly occurring during whimpers and screams. This contrasts our own predictions as well as what has been observed in the human infant literature. Human infants do not typically direct negatively valenced calls such as cries and screams to specific social partners using gaze (Oller, 2012). However, screams and whimpers are strong indicators of negative affect in young chimpanzees (Taylor et al., in prep), thereby questioning the previous claim that directed vocal behaviour presupposes flexible vocal production (Oller, 2012). For whimpers, which are mostly used during mother-offspring interactions (Plooij, 1984), this may reflect species differences in parenting behaviours between humans and chimpanzees. Humans typically practice alloparental care more so than chimpanzees (Newson, 2018), meaning it may be more important for young chimpanzees to identify a specific call recipient to elicit beneficial responses because care is mainly provided by specific individuals. Screams are usually used during agonistic interactions (Goodall, 1990). Third-party intervention in agonistic interactions is dependent on the characteristics of those involved (e.g. Romero & de Waal, 2010). Directing gaze towards another individual during screaming may provide information to others about who has been involved in a conflict. Further, given evidence of acoustically distinct scream types depending on an individuals' role in an agonistic interaction (i.e. victim or aggressor: Slocombe & Zuberbühler, 2005) directed screaming may also provide information about

who played what role in an agonistic interaction, thereby influencing when third-party responses occur.

The lack of an age effect on mutual gaze during communication is consistent with previous studies that reported mutual gaze occurred between mother-offspring dyads from early in chimpanzee ontogeny (Bard, 1994; Bard et al., 2005). Mutual gaze was significantly more likely to occur during screams, but mutual gaze during screams also seemed to reduce with age. That mutual gaze mostly occurred during screams contrasts the view that mutual gaze is typically associated with affiliative interactions (e.g., Feldman, 2012) and the expression of positive affect in infants (Colonnesi et al., 2012). However, mutual gaze primarily indicates engagement (Reddy, 2010), and while engagement might often occur during affiliative interactions, there is no reason to expect it is limited to them. Engagement provides a means by which individuals can understand others psychological states (Reddy, 2010, 2019). For chimpanzees, this might be particularly important in the agonistic interactions during which screams typically occur and wherein the risk of incurring physical injury and damaging social bonds is high (Goodall, 1990). However, since aggression is more common in older individuals (Anestis, 2006), this begs the question of why mutual gaze during screams significantly decreases during ontogeny. Screams are associated not only with agonistic interactions but also excitement (Clark & Wrangham, 1993) and are known to be used in combination with other call types. For instance, screams are incorporated into chimpanzee pant-hoots (Crockford & Boesch, 2005), which generally occur outside of dyadic social interactions and are largely produced by older males (Fedurek et al., 2013). In a recent study, we also showed that screams are more flexibly produced across affective states in older

individuals (Taylor et al., in prep). As such, the observed decrease in mutual gaze during screams in ontogeny may be related to an increase in the use of screams outside of the context of dyadic aggression.

We also observed that individuals of an older age used particular cues of directedness during vocal communication (i.e. body orientation, gaze direction, facial orientation, etc) more consistently (and therefore less flexibly). Those who were more consistent in how their vocalisations were directed were also more likely to elicit behavioural responses from social partners when they vocalised. However, there was no difference between call types in how flexibly vocalisations were directed towards social partners. Since our analysis included call types that are otherwise known to be flexibly produced (i.e. grunts, Dezechache et al., 2020; Taylor et al, in prep) and are therefore expected to be more freely directed (Oller, 2012), this finding contrasts both our own predictions, and research on human infants which shows flexibly produced speech-related vocalisations are more freely directed towards social partners than affectively-biased non-speech vocalisations such as laughs and cries (see Oller, 2012). The human communication literature indicates that during ontogeny consistent socio-communicative styles emerge such as ‘assertiveness’ or ‘responsiveness’, which are related to individual differences in personality traits and impact effectiveness in interpersonal interactions (Richmond & Martin, 1998). Such socio-communicative strategies are often effective in different circumstances, for example, assertive communicative styles are valuable in circumstances that require dominance (Richmond & Martin, 1998). This raises the question of whether consistent communicative styles in chimpanzees are related to an individual’s unique position in the social network, where positions are known to vary depending on factors

such as dominance and subordination (Funkhouser et al., 2018). This would be consistent with the observation of consistency increasing during ontogeny because infants mostly associate with their mothers and siblings, while juveniles begin to occupy their own distinct position in the social network (Goodall, 1990). To explore these possibilities, further work could examine the relationship between ontogeny, communicative style, and social network positions.

To conclude, in the present study we aimed to examine the ontogeny of directedness and engagement during chimpanzee vocal communication. We found that much like human infants, young chimpanzees routinely show directedness and engagement during vocal communication. Directedness appears to be a social competency that improves during ontogeny and enhances the effectiveness of vocalisations. As individuals became older, they also became more consistent in how they directed their vocalisations towards others, hinting at the possibility of individualistic socio-communicative styles in chimpanzees. Unlike human infants, however, directedness and engagement were mostly associated with calls that express negative affect, questioning the extent to which directedness and engagement presuppose flexibility in vocal production. Our findings therefore highlight both similarities and differences with patterns of directedness and engagement that are observed in human vocal ontogeny. However, since directedness and engagement are considered to be important turning points on the path towards language acquisition in human infants (e.g., Donellan et al., 2020), the present observations suggest that they may have offered a route towards language phylogenetically, as well as ontogenetically.

5. General Discussion

Overview

The overall aim of this thesis was to provide new insights into the evolutionary origins of language, by adopting a comparative-developmental approach. In particular, I aimed to evaluate the extent to which the infrastructural natural logic model (IFNL, see figure 1.1) (Oller, 2012), which describes ontogenetic changes in human infant vocal behaviour that underlie language development, also describes the process of chimpanzee vocal ontogeny. The IFNL model outlines both infraphonological changes (i.e. changes in signal structure complexity) and infrasemiotic changes (i.e. changes in signal meaning and function), that provide a foundation for the emergence of language in human ontogeny.

On the infraphonological side, preverbal human infants transition from vocal behaviour that is initially more acoustically graded, to a vocal repertoire comprised of a larger number of more acoustically discrete units (Stark, 1981; Oller, 2000, Vihman, 2014). In the first empirical chapter (chapter 2), I examined whether chimpanzees also show such ontogenetic changes in the size and structure of their vocal repertoire. Comparisons between chimpanzee infant and juvenile vocal acoustics showed that while there were no differences in vocal repertoire size, the infant vocal repertoire was significantly less acoustically graded compared to juveniles. This pattern contrasts what is observed in human infants, suggesting parallels between early vocal behaviour in chimpanzees and human infants may be related to vocal production patterns rather than acoustic changes.

To investigate further the potential parallels between human infant and chimpanzee vocal behaviour early in ontogeny, I evaluated whether young chimpanzees show signs of the infrasemiotic communicative capacities that define human infant vocal communication. Namely, I tested whether young chimpanzees show free expressivity and functional flexibility in the second empirical chapter. Free expressivity refers to the ability to express different affective states on different occasions with the same call type (Oller, 2012). Free expressivity is a component functional flexibility which is defined as the ability to achieve different functions on different occasions with the same call type by expressing different affective states with the same call type (Oller et al., 2013; Jhang & Oller, 2017). These are capacities that indicate a form of signal-function decoupling that provides an essential foundation for the emergence of language according to the IFNL model (Oller, 2012). Human infants show signs of free vocal expressivity and functional flexibility within the first 3 months of life (Oller et al., 2013; Jhang & Oller, 2017). Similarly, we observed free vocal expressivity in the grunts of infant and juvenile chimpanzees. However, evidence for functional flexibility of freely expressed calls was mainly observed among juveniles. This suggests that patterns of vocal production that are fundamental for the emergence of language may be rooted in our ape ancestry. However, human infants appear to harness this communicative competency to navigate social interactions much earlier in ontogeny than chimpanzees.

I built on the findings of the second empirical study (chapter 3) in the third and final empirical chapter (chapter 4). The IFNL model predicts that freely expressed call types should also be more likely to be ‘freely directed’ (i.e. calls can be directed towards specific social targets using behaviours such as gaze direction and facial orientation) (Oller, 2012).

It is also known more widely in the human infant literature, that directed vocal behaviour and engagement during vocal communication plays an important role in language development (Bruner 1973; Donnellan et al., 2020). Thus, in the final empirical chapter, we studied the ontogeny of directed vocal behaviour, engagement during vocal communication, and its impact on social interactions. In contrast to the IFNL model, freely expressed call types of chimpanzees (i.e. grunts) were not more likely to be directed towards social partners. In fact, affectively biased call types (i.e. screams and whimpers) were more likely to be directed towards social partners and occur during mutual engagement. With regard to ontogeny, directed vocal behaviour generally increased with age as is observed in human infants (Gros-Louis et al., 2014; Donnellan et al., 2020). However, there were different ontogenetic patterns in directedness depending on call type. Moreover, directed vocalisations were more likely to elicit behavioural responses from social partners, similar to human infants (D’Odorico et al., 1997). These findings suggest that directedness in chimpanzee vocalisations is indeed an effective communicative competency that improves during ontogeny. However, it was not predicted by the IFNL model, suggesting the capacity for directed vocal behaviour does not require a high degree of flexibility in vocal production.

Together, the studies that comprise this thesis suggest that the IFNL model does not provide an entirely accurate description of the *process* of chimpanzee vocal ontogeny. Nonetheless, several of the core capacities that provide a foundation for the emergence of language in human vocal ontogeny according the IFNL model (i.e. free expressivity, functional flexibility, and directedness) were observed in young chimpanzees, suggesting these capacities provide a foundation for language in phylogeny as well as ontogeny. This

implies that while the IFNL model does indeed highlight capacities that appear to be important for the emergence of language, it appears further work is needed to better understand the relationships between such capacities, which will in turn shed light on how these capacities collectively give rise to the emergence of language.

Main findings and implications

Infraphonology of chimpanzee vocal ontogeny

The ontogeny of language relies upon changes in signal structure as well as meaning and function (Oller, 2012). Mature language is commonly viewed as a system of discrete elements (i.e. phonemes) that can be combined recursively into higher order structures (words, sentences, etc) (Hockett, 1990; Zuidema & de Boer, 2009). Indeed, formal linguistic theory is built entirely on the assumption that language is composed of such units (Watumull et al., 2014). This property of language has been argued to be responsible for much of the complexity and diversity observed in languages. For example, according to the particulate principle of self-diversifying systems (Abler, 1989), systems made up of discrete units (i.e. molecules, genes, phonemes, etc) that can be combined into higher-order structures tend to be more diverse because combinations of discrete units often have properties that lie outside of the individual properties of constituents (i.e. in a sentence, the whole is greater than the sum of its parts), whereas blending of constituent units leads to combinations whose properties lie between two units. Thus, discrete units of sound are believed to be fundamental to language.

In the IFNL model, it is hypothesised that the raw material for constructing the discrete units that comprise language is rooted in vocal behaviour that is originally more

acoustically graded. This graded vocal behaviour is argued to be elaborated into a system of a wider range of more discrete vocal categories (Oller, 2012). Indeed, studies of early vocal ontogeny in human infants show that early vocal behaviour comprises a small number of acoustically graded vocal types which are subsequently elaborated into a larger repertoire of more discrete vocal types during what is widely known in the human vocal development literature as the ‘expansion stage’ of vocal development (Oller, 2000; Stark, 1981; Vihman, 2014). However, my acoustic study of chimpanzee vocal ontogeny showed precisely the opposite of this pattern – I observed no changes between infants and juveniles in the number of call types in the vocal repertoire, however, I did observe that the vocal repertoire became increasingly acoustically graded rather than discrete. These findings do not necessarily contradict the predictions of the IFNL model because the model does not predict communication systems cannot become increasingly graded. However, according to the IFNL model, ontogenetic changes in the size and structure of the chimpanzee vocal repertoire would not provide an adequate foundation for the emergence of language.

While my findings suggest human and chimpanzee vocal ontogeny are very different processes from an acoustic perspective, it is important to note that in phonetic studies of human infants, researchers typically exclude ‘non-speech’ sounds which are more affectively grounded such as cries and laughter (see Vihman, 2014). In the present study, I included many affectively biased call types (i.e. laughter, screams, whimpers and hoo calls). In humans, it is known that the production of affective vocalisations is more flexible later in ontogeny (see Pisanski et al., 2016). Ontogenetic changes in the acoustic variability of such vocalisations have seldom been studied in humans. However, a growing

body of evidence is revealing that affective vocalisations in human adults are more acoustically flexible than previously thought, showing systematic acoustic variation that depends on factors such as social partner (Bryant et al., 2018) and context (Szameitat et al., 2009), suggesting affectively biased vocalisations may show ontogenetic changes in acoustic complexity. As such, further studies adopting a similar approach to the present study focusing on affectively biased human vocalisations may reveal new parallels between human and chimpanzee vocal ontogeny.

Since chimpanzees show ontogenetic changes in vocal tract and laryngeal morphology that afford similar communicative possibilities to developing human infants (Nishimura et al., 2003, 2006), the question of why such different patterns are observed in human and chimpanzee vocal ontogeny remains. X-ray imaging studies of the macaque vocal tract show that the space of possible sounds that could in principle be produced is sufficient to support a language yet their vocal repertoire is comparably limited (Fitch et al., 2016), suggesting differences in neural control over vocal production rather than vocal tract morphology may explain species differences in vocal repertoire structure (Fitch, 2018). Chimpanzees that have learned to produce attention-getting sounds show increased neocortical grey matter volume in the ventrolateral prefrontal and dorsal premotor cortices (Bianchi et al., 2016) indicating some degree of flexibility in neural control over vocal production. However, directly comparative neuroanatomical studies of vocal control in chimpanzees and humans are sorely lacking. Consequently, the role of species differences in neural control over vocal production in explaining species differences in vocal ontogeny remains unclear.

The apparent difference in early changes in human and chimpanzee repertoire structure may be related to what those repertoires are designed to communicate. From an information theoretic perspective, the structure of a repertoire constrains what can possibly be communicated with that repertoire (Doyle et al., 2011). Words, for instance, are commonly believed to be referential (i.e. words are labels for discrete classes of objects in the world) and consequently words often have distinct meanings, rendering discrete units well-suited to expressing those meanings (Hockett, 1990). Indeed, Hockett (1990) argued gradedness in signals is particularly problematic when the number of meaningfully distinct vocal units in a repertoire increases. My observations of increased vocal gradedness in chimpanzee vocal ontogeny could therefore suggest that chimpanzee call meanings are continuous rather than discrete, and chimpanzees become progressively more adept at making effective use of this system throughout ontogeny. For example, whimper calls commonly grade into screams (Slocombe & Zuberbühler, 2010) and both indicate negative affect (Taylor et al. in prep). Thus, whimpers and screams may indicate different degrees or varieties of negative affect that grade into one another. Since more discrete communicative units are less likely to be misinterpreted by listeners (McCowan et al., 1999), this might also suggest that early in ontogeny, when chimpanzees are heavily dependent on caregivers (Plooij, 1984), unambiguous communication in the form of more discrete call meanings may confer a survival benefit, thereby explaining my observation that infant calls were more acoustically discrete.

If more discrete vocal units may be beneficial early in chimpanzee ontogeny, what might then be the benefit of increased vocal gradedness later in ontogeny? In contrast to early ontogeny which is spent mostly with the mother in chimpanzees (Plooij, 1984), older

individuals arguably have more complex social lives because they have social bonds with a wider range of non-kin group members (Goodall, 1990), although in late adulthood males spend their time with a smaller number of individuals with whom they share strong affiliative relationships (Rosati et al., 2020). An influential idea is that navigating complex social worlds also requires more complex communication (Freeberg et al., 2012). Further, increasingly graded vocal repertoires have a greater information-encoding potential (Fischer et al., 2017). Therefore, a more graded vocal repertoire later in ontogeny may afford more nuanced forms of communication that could allow individuals to navigate a wider range of social circumstances that vary in subtle ways and are seldom experienced by infants. More detailed studies of the social correlates of subtle variation in call acoustics throughout ontogeny could shed light on this possibility. However, one could equally argue that there is also a correlation between the complexity of social life and complexity in the communicative repertoire in human ontogeny (Tomasello & Gonzalez-Cabrera, 2017). It appears then, both human and chimpanzee social life becomes increasingly complex in ontogeny, which is accompanied by an expected increase in complexity of the vocal repertoire, but such complexity takes a different form in humans compared to chimpanzees – a larger repertoire of more discrete vocal units in humans (Oller, 2000; Stark, 1981; Vihman, 2014) and increased acoustic gradation between vocal units in chimpanzees. Since chimpanzee listeners are likely capable of inferring the epistemic states of their conspecifics (Kaminski et al., 2008), gradedness (or ambiguity in call meaning) may not be a problem insofar as listeners pragmatically enrich the signal with contextual elements (that may include conspecifics' epistemic states).

The difference in human and chimpanzee repertoire structure ontogeny may be related to the relevance of the benefits that a large repertoire of discrete units carries. Humans commonly arrange discrete communicative units into higher-order structures, which carries a wide range of communicative benefits including increased expressive capacity with a limited vocabulary (Verhoef et al., 2014) and information recovery (i.e. in a syntactical structure, when a unit is missing, the type of unit can be inferred using syntactical rules) (Fisher et al. 2010). However, chimpanzee call sequences have been reported to comprise no more than 4 different call types with no evidence of higher-order patterning (Crockford & Boesch, 2005), suggesting little benefit for a repertoire of discrete call types in chimpanzees. Additionally, both mathematical models and experimental studies on humans show that systems comprised of discrete communicative units can be socially learned more easily than graded signals (Zuidema & de Boer, 2009), providing a potential benefit to developing a repertoire comprised of discrete units. However, there is limited evidence of social learning in chimpanzee vocal behaviour. Acoustically, population-level differences in call structure have been observed (Crockford et al., 2004; Mitani et al., 1999; Mitani et al., 1992), although the extent to which social learning is responsible for this has been heavily contested (Fischer et al., 2015). While there are some indications of vocal learning in nonhuman primates (Marshall et al. 1999; Wich et al., 2009), it seems clear that there is a limitation in comparison to humans and some non-primate taxa (birds: Mason et al., 2017; sea mammals: Ravignani et al., 2016; bats: Prat, Taub, & Yovel, 2015), so that learnability of vocal structures appears to be of relatively smaller importance to chimpanzees.

Cognitive capacities have also been argued to constrain the structure of communication systems (Corballis, 2019). Experimental studies of auditory perception in humans and chimpanzees show that while both humans and chimpanzees are able to reliably categorise discrete sounds, chimpanzees show more errors in categorisation than humans (Kojima et al., 1989). Indeed, human cognitive capacities for categorical perception and hierarchical thought have been suggested as explanations of why language has the structure that it typically has (i.e. a system of discrete units recursively organised into higher-order structures) (Pinker & Jackendoff, 2005). Although, one should note that there also exist natural languages, such as the Pirahã language spoken by indigenous people of the Amazon, with no such structuring (Zwart, 2011), questioning the extent to which special-typical cognition shapes languages. Furthermore, the aforementioned study of auditory categorization (Kojima et al., 1989) focused on chimpanzee perception of human consonant sounds. Studies examining the extent to which chimpanzees perceive sounds from their own vocal repertoire categorically would provide a clearer insight into the extent to which chimpanzee auditory perception and cognition explains ontogenetic changes in their vocal repertoire structure. Nonetheless, the apparent difference between humans and chimpanzees in the structure of their vocal repertoires may in part reflect species differences in cognitive capacities. Interestingly, information-theoretic studies of acoustic signal structure show that species often believed to have relatively more complex cognitive capacities, such as wolves and dolphins, also produce vocal signals with greater structural complexity (Kerschenbaum et al., 2018). Perhaps then, increased complexity in chimpanzee vocal repertoire structure in ontogeny is accompanied by changes in complex auditory cognition.

Despite various possible explanations for the apparent difference in the ontogeny of vocal repertoire structure between humans and chimpanzees, it is important to note that while formal linguistics views language as a system fundamentally composed of discrete units, phoneticians have argued that natural speech shows a considerable amount of gradation between phonetic units (Port & Leary, 2005). Moreover, there are cultural differences in the discreteness of the underlying phonological system of a given language, with both vowel and consonant sounds having intermediate forms that grade into one another in many languages (Inuit consonants: Ulving, 1953; Uto-Aztecan consonants: Langacker, 1976; Irish-English vowels: Hickey, 1984). As such, the apparent difference between humans and chimpanzees in the ontogeny of the vocal repertoire may be overestimated. This may reflect a bias in the IFNL model, which has been largely validated by research on Western infants (e.g., Buder et al., 2008; Iyer et al., 2016; Jhang & Oller, 2017; Oller et al., 2013).

Methodological differences in the study of human and non-human primate vocal repertoire structure are also important to consider. Phoneticians studying human speech establish discrete units first on an auditory basis, assess the validity of categories based on acoustic, kinematic, and physiological data, and confirm these units through reliability testing (Shriberg & Lof, 1991). A similar approach (with the exception of validation with kinematic and physiological data) is common in the identification of possible units of non-human primate vocal repertoires, although it has been argued to mis-represent actual vocal repertoires and lead to inconsistencies across studies (Crockford, 2019). Data-driven approaches such as those adopted in this thesis to assess the ontogeny of vocal repertoire size and structure, have not been used to measure early ontogenetic changes

in the size and structure of the human vocal repertoire. As such, further comparative studies adopting more similar methods will clarify the similarities and differences between human and chimpanzee vocal ontogeny and in turn provide clearer insights into whether the patterns observed in chimpanzee vocal ontogeny could, in principle, provide an adequate foundation for the emergence of language. It is important that such studies also sample infants from a range of language environments that reflect the extent of natural variation in phonological systems in order to ensure apparent species differences are not confounded with the aforementioned human cultural variation in speech characteristics.

To summarise, the chimpanzee vocal repertoire becomes increasingly acoustically graded during ontogeny but the number of units in the repertoire does not change between the infant and juvenile period. This contrasts with human vocal ontogeny, wherein an increasingly large number of discrete vocal units emerges from vocal behaviour that is originally more acoustically graded. This difference may be explained by a lack of selective benefit for using more discrete signals among chimpanzees, and also cognitive differences of signallers and receivers at the species level that shape the structure of communication systems. Equally, the apparent difference could be explained by the inclusion of affectively biased calls in the present study in tandem with the exclusion of such vocalisations in human studies on this topic, highlighting the importance of studying the ontogeny of human non-verbal vocal communication from an acoustic perspective. The idea that language fundamentally comprises a repertoire of discrete vocal units itself may in fact be an assumption made by formal linguistics that does not necessarily accurately describe real-world speech and also overlooks cultural differences in language structures.

Methodological differences between phoneticians studying human speech and animal communication researchers may further obscure the similarities and differences in the ontogeny of vocal repertoire structure between humans and chimpanzees. As such, further studies of comparative communication should aim to develop shared methodologies that can clarify this issue. It will also be important to examine the social correlates of more subtle variation in call acoustics later in chimpanzee ontogeny to better understand how they make use of the new communicative possibilities that they are afforded by increased vocal gradation. In doing so, we will be able to more clearly ascertain what implications chimpanzee vocal ontogeny has for our understanding of the evolutionary origins of language.

Infrasemiotics of chimpanzee vocal ontogeny

Mainstream primatological research aimed at understanding the evolutionary origins of language has typically made an implicit commitment to the picture-theory of language whereby language is considered to be a system of labels that have a one-to-one correspondence with objects in the world. In the introduction, I discussed critical limitations of this theory of language, and argued for the need for primatological research to embrace alternative conceptualizations of language to shed new light on the evolutionary origins of language. In particular, I advocated the need to explore pragmatic linguistic theories whereby communication depends largely on what individuals *do* with their communicative behaviours, rather than only information content inside of signals which is typical of both the picture-theory of language and the code model of communication adopted by animal communication researchers more generally. Indeed, such capacities provide a foundation for language development in human ontogeny. In

chapters 3 and 4, I asked whether such capacities provide a foundation for language in phylogeny also, by examining whether these capacities are characteristic of early vocal behaviour in chimpanzees, and if so, to what extent is this similar to the patterns observed in early human vocal ontogeny.

Free expressivity & functional flexibility

According to the IFNL model, one of the most fundamental infrasemiotic capacities required for language development is the ability to express different meanings with the same communicative unit on different occasions, known as free expressivity (Oller, 2012). Freely expressed signals are also theorised to be functionally flexible, meaning the role a communicative unit plays in a social interaction depends on what that unit expresses on a particular occasion (Oller, 2012). Indeed, this is observed in early human vocal behaviour – from 3 months of age, infant protophones express positive, neutral and negative affective states on different occasions, whereas laughs and cries are strongly biased towards expressing positive and negative affective states respectively (Oller et al., 2013). Consistent with the predictions of the IFNL model, infant protophones are also functionally flexible – caregiver responses depend on what the protophone expresses on that occasion (Oller et al., 2013).

In chapter 3, I found that in both infant and juvenile chimpanzees, grunts were freely expressed (i.e. they expressed positive, neutral and negative affective states on different occasions and did not show any significant bias towards a particular affective state), whereas laughs, whimpers, screams, and hoo calls showed significant affective bias. Much like human infants, our data suggest that from early in ontogeny, the vocal repertoire of

young chimpanzees comprises a mixture of freely expressed and more affectively biased calls. Given evidence that chimpanzee grunts are produced across a range of contexts (Slocombe & Zuberbühler, 2010), the finding that grunts were also freely expressed is consistent with the predictions of the IFNL model, which posits that freely expressed calls must not be context limited. Further, by showing that this capacity is present in both infant and juvenile chimpanzees, this replicates and extends the findings of Dezechache et al. (2020), who showed free expressivity of infant chimpanzee grunts within the first year.

My findings of free vocal expressivity in young chimpanzees are also interesting within the context of the modality of language origins debate (i.e. gestural vs. vocal origins). Pollick and de Waal (2007) directly compared vocalisation and gesture production in chimpanzees and bonobos, finding that gestures were produced across a wider range of contexts, leading the authors to conclude that gesture is a more flexible modality in chimpanzees and therefore the most likely the modality within which language evolved. However, the authors focused only on two call types - screams and pant hoots. Screams in the present study were affectively biased, as were hoo calls, which are similar to elements within the pant-hoot call series (Goodall, 1990; Fedurek et al., 2013). Grunts, by contrast, were relatively more flexibly produced in my data. The present findings, with a fuller focus on the whole chimpanzee vocal repertoire, therefore suggest the belief that gestures are more flexible than vocalisations with regard to production may result from a narrow focus on a small number of call types, as well as reliance on a definition of flexibility that is not entirely comparable with the forms of flexibility shown in human infants (i.e. contextual flexibility rather than free expressivity).

In line with the predictions of the IFNL model (Oller, 2012; Oller et al., 2013), freely expressed calls (i.e. grunts) were the only call type that showed evidence of functional flexibility in this study. Together, our findings of free expressivity and functional flexibility suggest that characteristics of vocal production that are believed to provide an essential foundation for the emergence of language during ontogeny, also provide an essential foundation for the emergence of language in phylogeny. The observed patterns do, however, differ from the pattern observed among human infants in critical ways. Most notably, while free expressivity was observed in both infants and juveniles, evidence for functional flexibility was observed only among juveniles. This contrasts human infants, who showed both free expressivity and functional flexibility at 3 months of age (Oller et al., 2013). This does not contradict the IFNL model, since the model only predicts freely expressed call types *can* be functionally flexible, not that they *must* be. However, it is important to address because evolutionary animal communication theory posits that call types evolve only if they produce systematic effects in others (Maynard-Smith & Harper, 2003). The observation of functional flexibility among juveniles but not infants may indicate an improvement in socio-communicative competency during chimpanzee ontogeny, perhaps through learning. This interpretation is consistent with evidence that chimpanzees can learn to produce sounds to fulfil novel functions (Taglialatela et al., 2012), which arguably indicates improved social competency in vocal communication. However, infant social partners were also mostly mothers, whereas juveniles interact more with non-maternal group members, meaning the difference between infant and juvenile functional flexibility may be related to a difference social partner rather than communicative competency.

The observation of functional flexibility in the vocal behaviour of young chimpanzees shows for the first time that such an ability is present in non-human primates. However, the observed patterns of functional flexibility in chimpanzee juveniles were less pronounced than has been reported for human infants (i.e. *relative* differences in effects on social partner as a function of subject affect rather than *absolute*). Since human infants are more altricial than chimpanzees, having a longer period of dependency on caregivers wherein they depend on co-operation from caregivers for survival (Zemeloff & Boyce, 1982), this could reflect a species difference in social partner responsiveness. However, maternal responsiveness has been shown to be dependent on maternal attachment (Raval et al., 2001), which is known to vary cross-culturally (Carlson & Harwood, 2003). Therefore, the patterns of caregiver responsiveness observed by Oller et al. (2013), who focused on American families, may not be typical of human caregiving. Further cross-cultural studies may reveal patterns more similar to that observed in chapter 3. Methodological differences may also provide an explanation – human infants and their caregivers in Oller et al's (2013) study were observed in the lab, meaning demand characteristics may influence caregiver responsiveness. Indeed, comparisons of laboratory data and observational data collected in the home confirm that caregivers are more responsive to their infants in the lab than in the home (Belsky, 1980). Additionally, the laboratory environment likely removes most of the extraneous demands usually placed upon caregivers in everyday situations (Carrier et al., 2015), thereby affording them the ability to respond to infants more regularly and promptly.

While the findings were consistent with the IFNL model in many ways, the present observations also suggest a minor alteration to the model may be needed. In particular,

the IFNL model makes a categorical distinction between freely expressed and affectively biased calls (Oller, 2012; Oller et al., 2013). However, in chapter 3 I observed that affectively biased calls were not always biased towards a particular affective state. For example, juvenile whimpers were produced significantly differently across affective states, rarely expressing positive affect. However, there was no significant difference in neutral vs negative whimpers. By contrast, juvenile laughs consistently expressed positive affect. Free expressivity does not therefore appear to be a binary characteristic of a call type, but rather present to different degrees in different call types. Indeed, while human infant laughs and cries appear to be strongly affectively biased (Oller et al., 2013), human adults have more volitional control over such affectively biased vocalisations (Pisanski et al., 2016), indicating some degree of flexibility in vocal production. Since the IFNL model aims to outline the kinds of infrastructure that are believed to be essential for the emergence of language, this raises the question of what degree of freedom in vocal expressivity is required for language-like features of communication systems to emerge, and what do intermediary forms of flexibility afford signallers?

Directedness & engagement

According to the IFNL model, call types that are freely expressed are also expected to be ‘freely directed’, meaning individuals are able to direct their communicative acts towards social partners (i.e. by means of gaze directedness or facial orientation) (Oller, 2012). The rationale is that if calls are not freely expressed they are not under the individuals’ control, and if calls are not under the individuals’ control, they do not have the opportunity to direct communicative acts towards specific targets. Indeed, such a pattern is observed in human infants (Oller, 2012), which also appears to play an important role in

the development of communication and language. For example, infants who show more directed vocalisations are more likely to elicit responses from caregivers, suggesting directedness is a social competency that is important for effective communication (Gros-Louis et al., 2014). Additionally, infants who show more directed vocal behaviour within the first year of life show a greater expressive vocabulary in the second year (Donnellan et al., 2020), demonstrating a direct relationship to language development.

In chapter 4, we examined the relationship between age, call type, and directedness. We found that much like human infants, directedness showed a general increase in chimpanzees during ontogeny and calls that were directed towards social partners were more likely to elicit responses from them. Since directedness increased with age and increased the probability of eliciting a behavioural response from a social partner, this may help to explain why functional flexibility, which depends on social partner responses, was observed in juveniles but not infants. However, patterns of directedness across call types did not match the predictions of the IFNL model. Overall, screams and whimpers, which show a strong affective bias, were more directed towards social partners. Moreover, there was no difference between call types in the flexibility with which calls were directed towards social partners. This stands in direct contrast to the IFNL model which predicts grunts should be more directed because they are the only freely expressed call type in the vocal repertoire of young chimpanzees.

One explanation of this is that the predictions of the IFNL model with respect to directedness are built upon the implicit assumption that directed vocal behaviour indicates intentional communication (Bruner 1973; Donnellan et al., 2020) which requires

flexible vocal production, whereas affectively grounded vocal behaviour is automatic and involuntary (Scherer, Mortillaro, & Mehu, 2013). Such reasoning is indeed consistent with the IFNL model's prediction that freely expressed calls will also be freely directed whereas affectively biased calls will not. Although I am cautious about the interpretation of my data as providing evidence for intentionality, a growing body of literature does support the claim that chimpanzee vocalisations (including alarm calls as well as grunts) are intentional (e.g. Crockford et al., 2015; Schel et al., 2013). Consequently, the IFNL model may not have accurately predicted directedness in chimpanzee vocal behaviour because the model implicitly embraces a theoretical assumption (i.e. that affectively grounded communication is uncontrolled) that may not be entirely true. Since affectively biased calls were more directed in chapter 4, my findings therefore could question the distinction between intentional and affectively grounded communication. Indeed, at least in human adults, affectively valenced vocalisations are not entirely involuntary. Laughter, for example, can be volitionally produced and express a range of subtle varieties of positive affect (Scott et al., 2014), demonstrating that affectively grounded vocal behaviour is not entirely uncontrolled.

Despite the finding that the IFNL model did not accurately predict directedness, the results nonetheless could imply that this capacity may be a phylogenetic precursor to language. This interpretation must, however, be taken with caution. People with congenital visual impairments who cannot visually direct communicative acts manage to learn language, despite some limitations in pragmatic language use (Tadić et al., 2010). Moreover, cultures vary considerably in the extent to which they participate in such visually directed forms of communication (Fogel et al., 1988), yet almost all typically

developing individuals across cultures acquire language. Further conceptual work on directedness clarifying what it means to direct communicative acts in non-visual domains may resolve this issue. However, it may also be the case that while directed forms of communication *can* offer a path towards language, there are alternative routes available. In turn, this implies the IFNL model identifies capacities that *can* give rise to language, rather than capacities that one *must* have in order to develop language.

The role of directed communication in social interactions and language development is theorised to be mediated by the visual mutual engagement between social partners that directed communicative acts establish (Franklin et al., 2014; Lavelli & Fogel, 2005). In chapter 4, we found that visual mutual engagement routinely occurs during vocal communication, and changes with age depending on call type. Despite playing an important role in language development, the IFNL model entirely overlooks this aspect of directed vocal communication. The reason for this, is that the IFNL model focuses on individual capacities in isolation of the world around them. However, language development is at least in part an interactive process (Chapman, 2000). This suggests an adequate theory of the origins of language, either in ontogeny or phylogeny, must account for interactive aspects of communicative development, as well as the infrastructure required for individuals to possess in order to make such developmental processes possible.

Summary

In contrast to the infraphonological aspect of the IFNL model, several of the infrasemiotic capacities that the model posits provide an essential infrastructure for language appear

to be present in the vocal communication of young chimpanzees. Infant and juvenile chimpanzees show free vocal expressivity, and consistent with the IFNL model, freely expressed calls are also functionally flexible, although functional flexibility was only observed among juveniles. I also observed that vocal behaviour was routinely directed towards social partners and regularly occurred during visual mutual engagement.

Directedness also increased with age and increased the probability of elicited behavioural responses from a social partner, which may help to explain why functional flexibility was observed in juveniles but not infants. While directed vocal behaviour was routinely observed and showed some developmental similarities with directed communication in human infants, it was not predicted by the IFNL model. Overall, this suggests the IFNL model may require revision with regard to the relationship between infrasemiotic capacities such as free expressivity and free directedness. However, the presence of these capacities during early chimpanzee vocal communication and their impact on social interactions lends support to our argument that by embracing a pragmatic philosophical conceptualisation of language that focuses on what individuals *do* with their communicative behaviours, primatological research can shed new light on the evolutionary origins of language.

Infrastructural natural logic model & the evolution of language

In the introductory section, it was argued that a comparative-developmental approach is essential to further elucidate the evolutionary origins of language. To guide such an approach, a theoretical model that identifies the features of developing communication systems that provide a foundation for the emergence of language is essential. As one of the most comprehensive models available, I explored the extent to which Oller's (2012)

IFNL model described early vocal ontogeny in chimpanzees. Thus far, the extent to which particular infrasemiotic and infraphonological characteristics are present in the vocal behaviour of young chimpanzees has been discussed. I will now address the extent to which the model accurately describes chimpanzee vocal ontogeny overall and highlight general limitations of the approach adopted in this project.

On the infraphonological side, chimpanzee and human vocal ontogeny appear to be very different processes. While the acoustic complexity of vocal behaviour increases during ontogeny in both humans and chimpanzees, chimpanzee vocal ontogeny is a process whereby vocal types become progressively more acoustically graded. By contrast, human vocal ontogeny from an infraphonological perspective, is a process whereby an increasingly larger repertoire of more acoustically discrete sound types emerge from vocal behaviour that is originally more acoustically graded (Oller, 2000; Stark, 1981; Vihman, 2014). According to the IFNL model then, ontogenetic changes in chimpanzee repertoire structure would not provide an adequate foundation for the emergence of language. As such, the IFNL does not accurately describe the process of chimpanzee vocal ontogeny from an acoustic perspective. This does, however, identify a potentially key evolutionary transition in the hominin lineage that may have facilitated the emergence of language – the ability to create a large vocal repertoire of discrete elements using pre-existing acoustically diverse vocal behaviour. Indeed, such acoustically diverse vocal behaviour, as is apparently characteristic of chimpanzees, may be an important phylogenetic precursor to language as it is in ontogeny. Without such prior diversity, it is difficult explain where the raw material of language might originate. This possibility could be further explored by quantifying acoustic gradation in the vocal repertoire of other

great apes. Indeed, other great ape species are known to be able to produce acoustically diverse vocalisations (orangutans: Lameira et al., 2016; bonobos: de Waal, 1988).

However, if chimpanzees and bonobos show higher levels of acoustic gradation than other great ape species such as gorillas and orangutans, this could suggest high levels of acoustic gradation are an important phylogenetic precursor to language.

On the infrasemiotic side, the IFNL model identifies several capacities which appear to be characteristic of chimpanzee vocal behaviour including free expressivity, functional flexibility, and free directivity. Not only does the model identify capacities that are present, it also makes some accurate predictions about the relationships between those capacities. For example, the model predicts contextually flexible call types are the only call types that can be freely expressed, and that freely expressed call types may also be functionally flexible. Consistent with this, grunts, which have regularly been identified as a chimpanzee call type produced flexibly across contexts (Plooij, 1984; Goodall, 1990; Slocombe & Zuberbühler, 2010; Crockford, 2019), were the only call type that showed both free expressivity and functional flexibility. That the model makes accurate predictions about the relationships between capacities is notable, because this is a centrepiece of the underlying rationale of the model. While Oller (2012) embraced Hockett's (1960) 'design features' approach to comparative communication, it was argued that Hockett's (1960) approach was limited in that it did not acknowledge the relationship between the different design features. The IFNL model was an attempt to outline the relationship between design features of language in a linear way that describes the essential infrastructure required for the emergence of language, in ontogeny or phylogeny.

The IFNL did, however, also make unsupported predictions about the relationship between infrasemiotic capacities. In particular, the model predicts that freely expressed calls should be more freely directed, yet in chapter 4 I observed that affectively biased calls were more likely to be directed, and there was no difference between call types in how flexibly they were directed towards social partners. This apparently contrasts the literature on human infants, which does support the IFNL model in this respect (Oller, 2012). Since a sizable body of evidence supports the argument that directed vocal behaviour plays an important role in language development (Lavelli & Fogel, 2005; Gros-Louis et al., 2014; Donnellan et al., 2020), this suggests that the IFNL model may in some respects need revising regarding the relationship between some of the infrastructural capacities required for language to emerge.

According to the IFNL model, infrastructural capacities are ordered in the way that they are because each capacity is made possible by the preceding infrastructure. That is, in the jargon of the model, each capacity that emerges logically presumes the preceding capacities, and therefore capacities emerge according to a ‘natural logic’ (Oller, 2012). While such dependencies appear to occur for some language-relevant capacities (i.e., contextual flexibility, free expressivity, and functional flexibility), my observations on directedness suggest at least some language-relevant capacities do not show such dependencies. One interpretation of this pattern is that language is not a single phenotype that emerges in a linear fashion as the IFNL model assumes (Figure 5.1A). Rather, language appears to be a mosaic phenotype that emerges from an interaction between distinct capacities with their own developmental trajectories (Figure 5.1B) and

may therefore have evolved as such. The idea of mosaic evolution is widely accepted in evolutionary biology (Carroll, 1997), and it has also been suggested to apply to language evolution previously (Hurford, 2003). This possibility is further supported by comparative studies of neural architecture in chimpanzees and humans, which have shown that human brain evolution is a mosaic process (Gómez-Robles et al., 2014).

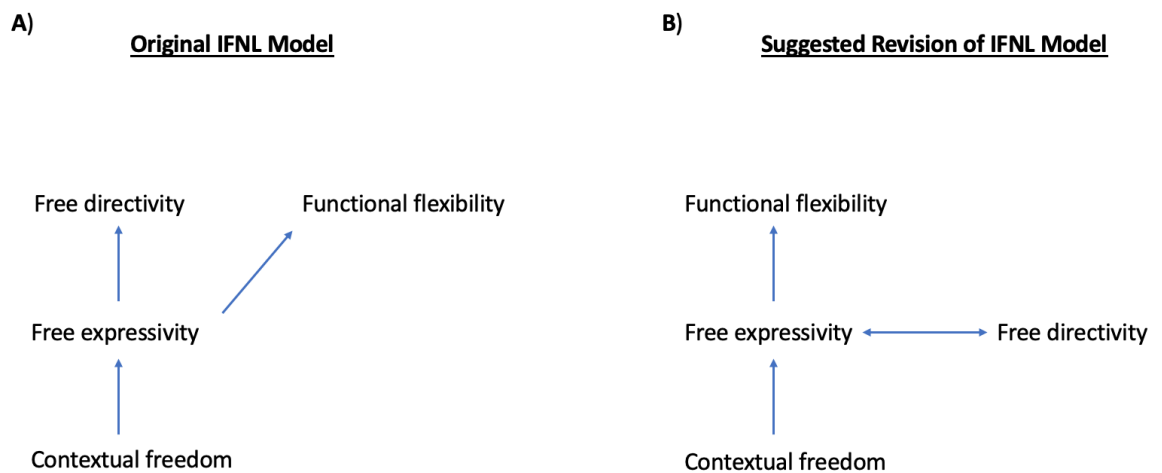


Figure 5.1 Original IFNL model wherein all language-relevant communicative capacities require prior infrastructure, adjusted from Oller (2012) (A). A suggested revision of the IFNL model wherein language emerges from the interaction between distinct capacities based on observations of chimpanzees, some of which develop as the original IFNL model predicts (B).

Despite the limitations on the predictions of the IFNL model, it is important to also consider general methodological limitations of the approach adopted in this thesis to understand why the predictions of the IFNL model may not have been entirely supported. Firstly, my sample differs considerable in age range from those samples used to validate

the IFNL model in human infants. Much of the research supporting the IFNL model in human infants focuses on infants up to 24 months of age (e.g., Oller et al., 2013; Jhang & Oller 2017; Donnellan et al., 2020). In the present sample, chimpanzee ages range from 0 – 10 years. While the adoption of a comparative-developmental approach to language evolution represents progress from previous studies that largely rely on adult samples, such a wide age range could have obscured subtle patterns that occur early in ontogeny. More detailed studies focusing on earlier vocal ontogeny (i.e. within the first 24 months) may reveal patterns that are indeed consistent with the predictions of the IFNL model. Secondly, the present thesis only included chimpanzees, which may limit the conclusions that can be made about language evolution. Previous phylogenetic studies of great ape vocal communication (e.g. laughter: Davila-Ross et al., 2009) show that direct comparisons between humans and chimpanzees can be misleading due to species-specificity in traits and reversals of traits to ancestral states. As such, further studies with a wider range of species may indeed find further support for the predictions of the IFNL model and strengthen the conclusions that can be made about language evolution. Finally, one might argue that the extent to which chimpanzee vocal behaviour is comparable to human vocal behaviour is questionable, since chimpanzee vocalisations are non-verbal, while early human vocal communication comprises both verbal and non-verbal aspects. This may undermine the degree to which conclusions can be made regarding the IFNL model from observations of non-human primate vocal behaviour, suggesting the conclusions of this project should be taken with caution. However, the aspects of the IFNL model that were of focus in the present project were all characteristics of pre-verbal vocal communication in human infants, supporting the relevance of such a comparison in this case.

Overall, the IFNL model does not appear to be a fully accurate model of chimpanzee vocal ontogeny. Infraphonologically, the predictions of the model were not met.

Infrasemiotically, some predictions were supported but others were not. However, since many of the capacities outlined in the IFNL model were found in vocal communication among young chimpanzees, further studies looking for other capacities outlined by the model in chimpanzees may still be worthwhile. The observation that many of the capacities outlined by the IFNL model are present in both humans and chimpanzees, but the relationship between those capacities was not supported by chimpanzee data, may suggest that language is a mosaic phenotype that emerges from an interaction between different communicative capacities in both ontogeny and phylogeny.

General Strengths of the IFNL model

In adopting the IFNL model to guide a comparative-developmental enquiry into the evolutionary origins of language this thesis benefitted in a number of ways. Firstly, as discussed in the introductory section, primatology is often ambiguous about the philosophical conceptualisation of language to which it is committed. The problem that this introduces into the literature is that it obscures our understanding of the evolutionary origins of language because the extent to which language appears to be rooted in our ape ancestry largely depends upon the type of system that we consider language to be in the first place. By adopting a model with a clear commitment to a pragmatic conceptualization of language, more specifically an Austinian conceptualization of language, I am able to infer, given my results, that *if* we take language to be a fundamentally pragmatic system in the Austinian sense, *then* the foundations for such

communication may be rooted in our ape ancestry. Further studies could in turn build upon this work either by further exploring the Austinian framework in non-human primates, or by making explicit contrasts with other comparable philosophical frameworks such as the Gricean pragmatic conceptualization of language. These efforts could contribute to resolving debates surround the type of system language actually is.

The IFNL model brought clarity to this thesis not only through its philosophical roots, but also by identifying clear and novel targets for analysis. Previously, primatological research has largely focused on the presence of mature features of language in ape communication systems. Most vividly, great ape language studies have shown apes are capable of learning to use symbols to communicate and even show evidence of acquiring aspects of the generative structure of language in their use of such symbols (orangutans: Miles, 1990; bonobos: Savage-Rumbaugh et al., 1993; gorillas: Patterson & Cohn, 1990; chimpanzees: Gardner, Gardner & Van Cantfort, 1989;). Notably, these successes have been largely limited to the gestural domain, with attempts to teach vocal communication skills being comparatively limited in success (e.g. Hayes & Hayes, 1952), although some more recent studies do appear to show some degree of vocal learning of novel sounds in great apes (see Wich et al., 2009, 2012; Lameira et al., 2015, 2016). Further, these complex language-related abilities in trained apes appear to be comparatively limited among wild populations (e.g. Boesch, 1991; Crockford & Boesch, 2005) which are arguably more informative regarding the phylogenetic history of language (MacDonald & Ritvo, 2016). In human ontogeny, these capacities are rooted in developmentally prior capacities that make features of mature language possible (Oller, 2000, 2012).

Consequently, by focusing on mature features of language, many of these prior capacities

may be overlooked in non-human species. This places a critical limit on our understanding of the evolutionary roots of language. With the new focus that follows from the IFNL model, the results of this thesis revealed the presence of important precursors to mature language features in chimpanzee vocal communication. Since few studies have searched for evidence of such capacities thus far, the IFNL model provides fertile grounds for further exploration in other primate species and perhaps even in more distantly related mammalian species.

The possibility of exploring the IFNL model further in other species, and indeed chimpanzees in the case of the present thesis, is enabled by its root in Hockett's (1960) notion of 'design features'. In the IFNL model, capacities are features that a system would require in order to implement a language. Since capacities in the IFNL are design features rather than concrete features of language, the capacities specified by the model can be searched for and even compared across species. The applicability of this model across species affords the benefit of confidence in evolutionary inferences. Often, the same concept may be used differently in different disciplines. For example, human words can be considered to refer to objects in the world and this can be confirmed through self-report methods (Paulhus & Vazire, 2007). Such methods cannot be used for animal communication researchers, who instead study 'functionally' referential calls (e.g. Slocombe & Zuberbühler, 2005). Using different conceptualisations of the same phenomenon undermines the evolutionary inferences that can be made because it is not entirely clear how comparable these different conceptualisations are (e.g. Boesch, 2007). My results in this thesis are, by contrast, closely comparable with the research conducted on human infants within the framework. Consequently, I am able to conclude that if these

capacities are important for language, and we are willing to attribute them to human infants, then we ought to be willing to attribute them to young chimpanzees also.

While the capacities in the IFNL model are conceptualised at the level of Hockett's (1960) design features, the model was also developed in response to a criticism of Hockett's (1960) approach. More specifically, Oller (2012) argue that Hockett (1960) did not specify the relationship between design features and could therefore not explain the emergence of language, either in ontogeny or phylogeny. By specifying the relationship between capacities, the IFNL model offers a rich source of predictions. Indeed, the hypotheses of this thesis were largely derived from this aspect of the IFNL model and further studies may also benefit greatly from this property of the framework. While the predictions were not entirely supported, this approach offers the possibility of integrating many aspects of language that are usually studied independently into a single coherent framework. In turn, this can enrich our understanding of the evolutionary origins of language by providing a framework that allows us to put the different pieces (i.e. language-related capacities) of the puzzle together.

General Weaknesses of the IFNL model

Given the range of aforementioned strengths of the IFNL model, one might ask why the model has not been utilised more broadly in the comparative, and indeed developmental literature. Herein lies a key weakness of the model – it is heavily laden with jargon terms that have often been misunderstood and described elsewhere under different names. For example, contextual freedom, the foundation upon which the model is built, can understandably be interpreted as referring to communication that is free (i.e. outside of)

any context. On this interpretation, scepticism about the foundations of the model may arise because real world communication is always embedded in some form of context (Jones, 2013). This interpretation is not, in fact, the intention of Oller and colleagues (personal communication), who instead used the term to refer to contextual flexibility and have since refined their terminology as such. Science is built around a shared lexicon, and therefore terminological misinterpretations of this sort represent a barrier to wider exploration of the model across species.

The IFNL model may be unappealing to some not only due to terminological challenges. Another key weakness of the model is that it rests heavily on a small set of fundamental assumptions that may have serious implications for subsequent work built upon these foundations. More specifically, according to the IFNL model, the entire infrastructure of early vocal development, both infraphonologically and infrasemiotically, is built upon a single foundation – contextual freedom (or, contextual flexibility). One unresolved challenge of implementing this model is deciding upon clear criteria for deciding when contextual freedom is present, or what degree of freedom is required for the development of subsequent capacities. Although studies have convincingly shown chimpanzees are capable of vocalising in a flexible and voluntary manner (Davila-Ross et al., 2011; Hopkins, Taglialatela, & Leavens, 2011; Crockford et al., 2012; Schel et al., 2013; Crockford et al., 2018), the presence of this ability, contextual flexibility, in non-human primates and its relation to the form of contextually free vocal behaviour seen in human infants is an open question (see Seyfarth & Cheney, 2018). The answer to this open question will have important implications for the significance of the findings in the present thesis in relation to the evolutionary origins of language. For example, if

chimpanzee calls are not contextually flexible in a manner deemed relevant for the IFNL model, this would in turn question the relevance of data showing free expressivity and functional flexibility in chimpanzee vocal ontogeny. Going forward, this will be an important issue to consider in applying the IFNL model to non-human vocal communication with a view to gaining new insights into the evolutionary origins of language.

The aforementioned uncertainty regarding the degree of contextual freedom required for the emergence of language-like features of a communication system highlights another weakness of the IFNL model – the categorical nature of the capacities specified within it. This is particularly problematic when it comes to the evolutionary origins of language. Evolutionary similarities and differences between species are often differences of *degree* rather than *kind* (West-Eberhard, 2005). As the IFNL model specifies language-relevant vocal capacities categorically, it precludes the possibility of identifying similarities in the degree to which a given capacity is present. In turn, this may mask important evolutionary precursors to language-relevant vocal capacities. This limitation of the IFNL model could be remedied methodologically. The IFNL model *conceptualises* capacities in an implicitly binary manner, but the model offers few methodological techniques to operationalise these capacities. Of course, this is not unique to the IFNL model, as most models do not explicitly offer methodological recommendations. Nonetheless, further work could develop methodological techniques that operationalise capacities as continuous variables. Indeed, such an approach was taken by Dezechache et al. (2020) who measured the degree of affective biasing in infant chimpanzee call types, although direct comparisons between species are yet to be conducted.

The IFNL model was developed in response to a crucial limitation of Hockett's (1960) work, namely, that Hockett (1960) did not specify the relationship between communicative capacities deemed fundamental for language. While the IFNL model does specify the relationship between capacities within branches of the model (i.e. infrasemiotic and infraphonological) which represents a clear improvement on previous work, the model does not specify the relationship among capacities *between* these branches. This is not only important for the theoretical coherence of the overall model, but also leads to interpretational issues of the results of studies guided by the IFNL model. For example, Dezechache et al (2020) found that while infant chimpanzee grunts were routinely produced across positive, neutral, and negative affective states, grunts were also acoustically distinct according to affective state. One interpretation of these findings is that infant grunts are freely expressed and contain prosodic acoustic variation according to affective states, much like the same word expressed in different ways would show systematic acoustic differences (Zec, 2005). However, another interpretation is that grunts are not freely expressed, and instead comprise a set of distinct sub-types that are fixed in the way that they are expressed. The IFNL model in its current form offers no conceptual tools for dealing with such issues because it does not specify the relationship between infrasemiotic and infraphonological capacities.

Broader applicability of the IFNL model

Having discussed the general strengths and weaknesses of the IFNL model, I will now turn my attention to the broader applicability of this model across species. In the present thesis, I aimed to adopt methods closely aligned with those used in the human infant

literature with the aim of ensuring comparability of the findings across species and in turn bolstering evolutionary inferences. Given the close evolutionary relationship between humans and chimpanzees (Prado-Martinez et al., 2013), this is arguably a justifiable approach. However, this approach becomes progressively less justifiable, and perhaps increasingly procrustean, as the evolutionary distance because a given species of interest and humans becomes greater. For example, directedness is typically operationalised using visual cues, but how could directedness be evaluated in other sensory domains? This may limit generalizability not only across species but also within species. Wild chimpanzees, for example, inhabit different ecological niches, including open savannah and dense forest, that vary considerably in visibility (Moore et al., 2017). Interestingly, all successful attempts to teach language-like communication to great apes have all occurred in environments with high levels of visibility between communicating dyads (see Krause & Beran, 2020). In this thesis, the sample occupied an environment with clear visibility and communication was mostly close-range and dyadic. However, in populations that occupy dense forest, long-distance communication may be more common. This may limit generalizability to other populations which in turn questions whether species-level inferences can be made in this case. It will be important to consider the sensory constraints on communication in different species and populations within species in future studies applying the IFNL model.

Another limiting factor on the broader applicability of the IFNL model is that it is primarily applied at the level of the call *type*. In human infants, for instance, it is protophones that are contextually flexible, freely expressed, functionally flexible, and show signs of directedness, while other call types such as cries and laughter do not (Oller 2012).

However, for many taxa, such as whales and songbirds, calls are seldom used in isolation, and instead, meaningful communication is achieved through a combination of calls (Engesser & Townsend, 2019). Species that adopt such communicative styles are widely considered to be useful models for understanding the syntactic aspect of language evolution (Suzuki & Zuberbühler, 2019). However, due to its focus on call types, the IFNL model would not pick out such communication systems as being relevant to the understanding of language evolution. In the present case, this is unlikely to be problematic because chimpanzee call combinations are comparatively limited (Crockford & Boesch, 2005). Further studies with other species should, however, consider this issue. One possible approach to make progress in this regard could be to treat meaningful combinations as the relevant unit of analysis rather than singular call elements.

The broader applicability of the IFNL model may also be limited not only by differences between the species or populations it may be interesting to apply it to, but also by the populations on which the model itself has been primarily verified. While the model has been shown to accurately describe early vocal behaviour in human infants (Oller, 2000; Oller, 2012; Oller et al., 2013), the majority of such work has been conducted on infants from Western populations, particularly in the USA (e.g. Buder et al., 2008; Iyer et al., 2016; Jhang & Oller, 2017; Oller et al., 2013). Thus, despite being inspired by Hockett's (1960) design feature approach, which aims to be comparable across species by describing the *type* of qualities a system must have in order to be able to implement a language rather than outlining *specific* qualities (for example, the ability to create symbols rather than the presence of a specific symbol), the IFNL model may in fact only describe qualities *sufficient* for some languages to emerge rather than qualities that are

fundamentally *necessary* for all languages. Further exploration of this model across cultures would help to clarify the extent to which the model is suitable for cross-species comparisons of communication and ascertain what such comparisons actually reveal (i.e. differences at the level of the species or the populations).

Methodological reflections

In evaluating the work in the present thesis, it is important to consider some of the key aspects of my methodological designs that permeate the empirical chapters and evaluate their impact on the general outcomes of this thesis. There are three issues in particular that are common to all the studies presented in this thesis. Firstly, the approach to ontogenetic stage classification. Secondly, the selection of a sanctuary sample. Finally, the approach adopted to call classification.

The main aim of the present thesis was to evaluate the process of chimpanzee vocal ontogeny. Ontogeny, broadly defined, refers to change over time within the lifetime of an organism (Gould, 1977). Ontogenetic time can be conceptualised in different ways (Brainerd, 1978). One approach is simply to examine changes over linear time (i.e. age), which we might call the ‘age approach’ (see Durkin, 1995). An alternative is what we could call the ‘stage approach’. The stage approach is informed by life history theory and involves dividing the lifespan into meaningful stages, often defined by particular morphological and behavioural characteristics (Locke & Bogin, 2006). The approach that is adopted can indeed impact the results of this thesis and is therefore important to consider. For example, adopting the stage approach forces subjects from a potentially wide age range into a single category, with the consequence of masking potentially

meaningful variation within the data. The stage approach may exceptionally be problematic in cases where ontogeny is a gradual process (i.e. small continuous changes rather than jumps between discrete characteristics). For example, chimpanzee infants in this thesis were defined as those who were aged 0-4, while juveniles were defined as those who were aged 4-10. This assumes within-category homogeneity and between-category heterogeneity (i.e. a 3 year old is more similar to a 1 year old than a 4 year old). Neither approach is necessarily correct and either can arguably be used depending on the aims and topic of a particular study.

In the present thesis, different approaches were adopted in different studies. In studies 1 (chapter 2) and 2 (chapter 3), the stage approach was adopted, while in study 3 (chapter 4) the age approach was adopted. The age approach, I argue, is suitable when a single individual-level dependent variable is used because age is also a single individual-level variable. In studies 1 and 2, there was no single dependent variable. In study 1, I aimed to study the size and structure of the vocal repertoire at different stages of ontogeny. The size and structure of the vocal repertoire is a complex target that is evaluated by aggregating data from many individuals and therefore cannot be reduced to a single individual-level measurement. Similarly, in study 2, I examined patterns of vocal production across affective states, which also cannot be reduced to a single individual-level measurement. As such, the stage approach was adopted in studies 1 and 2. By contrast, study 3 examined a single dependent variable, directedness, meaning the age approach could be adopted. Moreover, previous studies on the same topic as study 3 have shown that directedness in chimpanzee ontogeny is better described by age than stage (Laporte & Zuberbühler, 2011). Indeed, it would be possible to perform the analyses

of studies 1 and 2 at each age instead of each stage (i.e. to treat age as if it were a stage). However, this would require an enormous increase in the number of subjects and calls, as for some ages there was only a single subject in my dataset (i.e. Nancy, age 1 year). Further, this approach may be extremely difficult to interpret if patterns of statistical significance are not consistent across ages in a linear manner.

Another methodological choice that may impact results throughout the empirical studies of this thesis is the sample that was chosen. In this thesis, I decided to study a sanctuary population. While the site at which data collection takes place is often influenced by the sites at which researchers at a given lab have long term collaborations and indeed this is the case here, I would argue that a sanctuary population was well suited to the present study. Many of the communicative cues studied in this thesis were visual, for example - facial expressions, body postures, and gaze. Attempts to study these communicative cues in combination with vocal behaviour in wild chimpanzees has been previously limited by the poor visibility in wild populations, particularly those that occupy dense forest habitats (e.g. Laporte & Zuberbühler, 2011; Dezecache et al., 2020). By contrast, while Chimfunshi Wildlife Orphanage contains dense miombo forest, all recordings were taken in relatively large and open spaces near the perimeter of the enclosures. As such, visibility was very good in the present study, allowing highly detailed coding of behaviours such as facial expressions, body postures, and gaze. This was crucial to improving upon previous work, for instance, by allowing me to code a wider range of affective cues than previous studies on wild samples were able to (e.g. Dezecache et al., 2020) and ensuring comparability to the human infant literature wherein behaviours can also be coded in a high level of detail thanks to good visibility of subjects (e.g. Oller et al., 2013).

The choice of a sanctuary population was not only methodologically convenient given my aims but was also theoretically important. For example, Marler (1967) theorised that vocal signals will be more acoustically discrete in closed habitats with poor visibility because such signals can more easily be recognised, whereas vocal signals will be more acoustically graded in open habitats with good visibility because the benefits of increased information encoding potential in acoustically graded signals are not offset by discriminability since signals can be disambiguated using visual information. Since all subjects in the present thesis were reared in an open habitat with good levels of visibility, the study population provides a useful model for evaluating ontogenetic changes in vocal repertoire structure and the use of visual cues during vocal communication. However, further studies should also perform similar analyses in populations occupying dense forest habitats to further test Marler's (1967) predictions, and in turn establish whether the observed patterns are species-typical which may have important implications for the evolutionary origins of language.

While the habitat of the sample in this thesis may have impacted the results with respect to repertoire structure, the results may also have been impacted by the approach adopted to code the infant and juvenile chimpanzee vocal repertoire. The chimpanzee vocal repertoire is commonly described at two different levels – the call *type* and the call *sub-type*, the former being a broader category of calls that may encompass different varieties of the latter. For example, within the call type 'grunts', many report distinct sub-types such as food grunts, pant grunts, and soft grunts (Goodall, 1990; Crockford, 2019). In this thesis, I described the vocal repertoire at the level of the call type rather than sub-

type. This could have an impact on how the process of chimpanzee vocal ontogeny is understood. For example, if coding was performed at the level of the sub-type, then it could show that there are ontogenetic differences in repertoire size which were not observed in chapter 2 of the present thesis. Although, it is important to note that the aim of chapter 2 was not to identify the true repertoire size, but rather, to evaluate possible differences in repertoire size. Further, if distinct sub-types were present, they could have been extracted as a distinct cluster by the unsupervised learning algorithm, yet they were not. The primary benefit of the chosen approach to call coding is that vocal units can be directly compared at different ontogenetic stages since there is widespread agreement about the presence of different call types throughout ontogeny (e.g. Plooij, 1984; Kojima, 2008), whereas the presence of distinct sub-types is less clear due to a lack of empirical study (see Dezechache et al., 2020). An additional benefit is comparability with the human infant literature on the topic of the present thesis, wherein analyses are also conducted at the level of the call type rather than sub-type (e.g. Oller et al., 2013; Yoo et al., 2018), which was particularly important given the overarching aims of this thesis – to further elucidate the evolutionary origins of language. The inter-rater reliability testing performed in this thesis stands testimony to the discriminability of the coded call types and the validity of these call type distinctions was further supported by unsupervised machine learning analysis on an acoustic dataset which showed the coded call types were spontaneously separated into distinct clusters. The coded call types in this thesis therefore represent coherent units for analysis. An important future step will be, however, to examine when in ontogeny distinct call sub-types emerge.

To summarise, the present thesis was centred around 3 key methodological design choices – the approach to developmental stage classification, the choice of a sanctuary population, and the approach to coding the calls of the vocal repertoire. Each of these decisions, I argue, were justified given then aims of the studies that comprise this thesis. However, each could have plausibly had an impact on the results of this thesis. Further research, adopting different approaches may be informative about the impact of these decisions and could further present understanding of the process of chimpanzee vocal ontogeny and its relation to language origins.

Overall conclusions and future directions

The overall aim of this thesis was to provide new insights into the evolutionary origins of language by adopting a comparative-developmental approach. To pursue this goal, I examined the extent to which Oller's (2012) infrastructural natural logic model of language, which describes early vocal ontogeny in human infants, also describes early vocal ontogeny in chimpanzees. I found that from an acoustic perspective, human and chimpanzee vocal ontogeny appear rather different. Acoustic complexity in the vocal repertoire increases during human and chimpanzee vocal ontogeny, but the form of that complexity differs in ways that the IFNL model posits are crucial for the emergence of language. By contrast, chimpanzee vocal ontogeny was characterised by a variety of pragmatic communicative competencies that are also known to play an important role in language development. Pragmatics has long been known to be essential for a full understanding of language. Together, these findings suggest that language is a fundamentally pragmatic system in phylogeny, as well as ontogeny. By testing for such capacities in other primate species, future studies can reveal whether these capacities are

deeply rooted in our primate ancestry or are instead more evolutionarily recent abilities. Such studies are likely to be both theoretically and methodologically challenging. In the present case, methodologies from the study of human infant communication were adopted. However, such methodologies may not be applicable to other species that communicate using different modalities. Nonetheless, by revealing whether the abilities that provide an essential foundation for language are limited to great apes, or are more widely shared among primates and perhaps other mammals, we may be able to make considerable progress in understanding how it came to be the case that almost all living things communicate yet only humans have language.

References

- Aaron, P. G., & Joshi, R. M. (2006). Written language is as natural as spoken language: A biolinguistic perspective. *Reading Psychology, 27*(4), 263-311.
- Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures, 12*(1), 1-13.
- Alarcao, S. M., & Fonseca, M. J. (2018). Identifying emotions in images from valence and arousal ratings. *Multimedia Tools and Applications, 77*(13), 17413-17435.
- Anestis, S. F. (2006). Testosterone in juvenile and adolescent male chimpanzees (*Pan troglodytes*): Effects of dominance rank, aggression, and behavioral style. *American Journal of Physical Anthropology, 130*(4), 536-545.
- Arnold, K., & Bar-On, D. (2020). Primate pragmatics, expressive behavior, and the evolution of language. *Animal Behavior and Cognition, 7*(2), 117-130.
- Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language, 120*(3), 303-309.
- Aureli, F., & Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. *Behavioral ecology and sociobiology, 73*(1), 6.
- Austin, J. L. (1975). *How to Do Things with Words*. Clarendon Press.
- Bădescu, I., Watts, D. P., Katzenberg, M. A., & Sellen, D. W. (2016). Allop parenting is associated with reduced maternal lactation effort and faster weaning in wild chimpanzees. *Royal Society Open Science, 3*(11), 160577.
- Baker, K. C., & Aureli, F. (1997). Behavioural Indicators of Anxiety: An Empirical Test in Chimpanzees. *Behaviour, 134*(13-14), 1031-1050.
- Bard, K. A. (1994). Evolutionary roots of intuitive parenting: Maternal competence in chimpanzees. *Early Development and Parenting, 3*(1), 19-28.
- Bard, K. A. (1998). Social-experiential contributions to imitation and emotion in chimpanzees. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny*, (pp. 208-227). Cambridge University Press.
- Bard, K. A., Gaspar, A. D., & Vick, S. J. (2011). Chimpanzee faces under the magnifying glass: Emerging methods reveal cross-species similarities and individuality. In A. Weiss, J. E. King, & L. Murray (Eds.), *Personality and temperament in nonhuman primates* (pp. 193-231). Springer.
- Bard, K. A., & Leavens, D. A. (2014). The importance of development for comparative primatology. *Annual Review of Anthropology, 43*, 1-27.
- Bard, K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A., & Matsuzawa, T. (2005). Group differences in the mutual gaze of chimpanzees (*Pan troglodytes*). *Developmental Psychology, 41*(4), 616-624.
- Barton, K., & Barton, M. K. (2015). Package 'MuMIn'. *Version, 1*, 18.

- Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, 2(1), 74.
- Beck, B. (2010). Chimpanzee orphans: sanctuaries, reintroduction, and cognition. In E. Lonsdorf., S. Ross., & T. Matsuzawa (Eds), *The mind of the chimpanzee: Ecological and experimental perspectives* (pp. 332-346). University of Chicago Press.
- Belsky, J. (1980). Mother-Infant Interaction at Home and in the Laboratory: A Comparative Study. *The Journal of Genetic Psychology*, 137(1), 37–47.
- Bezdek, J. C. (1973). Cluster Validity with Fuzzy Sets. *Journal of Cybernetics*, 3(3), 58–73.
- Bianchi, S., Reyes, L. D., Hopkins, W. D., Taglialatela, J. P., & Sherwood, C. C. (2016). Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. *Scientific reports*, 6, 34733.
- Bloom, K., D’Odorico, L., & Beaumont, S. (1993). Adult preferences for syllabic vocalizations: Generalizations to parity and native language. *Infant Behavior and Development*, 16(1), 109–120.
- Bloom, K., & Lo, E. (1990). Adult perceptions of vocalizing infants. *Infant Behavior and Development*, 13(2), 209–219.
- Boesch, C. (1991). Symbolic communication in wild chimpanzees?. *Human Evolution*, 6(1), 81-89.
- Boesch, C. (2007). What makes us human (Homo sapiens)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology*, 121(3), 227.
- Bohn, M., Call, J., & Tomasello, M. (2015). Communication about absent entities in great apes and human infants. *Cognition*, 145, 63–72.
- Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior and Development*, 22(1), 65-85.
- Bradbury, J. & Vehrencamp, S. (2011). *Principles of Animal Communication* (2 edition). Sinaeur.
- Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences*, 1(2), 173-182.
- Brigandt, I., & Love, A. C. (2010). Evolutionary Novelty and the Evo-Devo Synthesis: Field Notes. *Evolutionary Biology*, 37(2), 93–99.
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1-22.
- Bruner, J. S. (1973). Organization of Early Skilled Action. *Child Development*, 44(1), 1–11.
- Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Amir, D., Chávez, B., ... & Fux, M. (2018). The perception of spontaneous and volitional laughter across 21 societies. *Psychological science*, 29(9), 1515-1525.
- Buder, E. H., Chorna, L. B., Oller, D. K., & Robinson, R. B. (2008). Vibratory Regime Classification of Infant Phonation. *Journal of Voice*, 22(5), 553–564.

- Carlson, V. J., & Harwood, R. L. (2003). Attachment, culture, and the caregiving system: The cultural patterning of everyday experiences among Anglo and Puerto Rican mother–infant pairs. *Infant Mental Health Journal*, 24(1), 53–73.
- Carrier, L. M., Rosen, L. D., Cheever, N. A., & Lim, A. F. (2015). Causes, effects, and practicalities of everyday multitasking. *Developmental Review*, 35, 64–78.
- Carroll R.L. 1997. *Patterns and processes of vertebrate evolution*. Cambridge University Press
- Cecilione, J. L., Rappaport, L. M., Verhulst, B., Carney, D. M., Blair, R. J. R., Brotman, M. A., ... & Hettema, J. M. (2017). Test–retest reliability of the facial expression labeling task. *Psychological assessment*, 29(12), 1537-1542.
- Chapman, R. S. (2000). Children’s Language Learning: An Interactionist Perspective. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(1), 33–54.
- Clark A. 1991. *The socio-ecology of wild chimpanzees vocal behavior in the Kibale Forest, Uganda*. (Doctoral thesis, University of Michigan, Michigan, USA).
- Clark, A. P., & Wrangham, R. W. (1993). Acoustic analysis of wild chimpanzee pant hoots: Do Kibale Forest chimpanzees have an acoustically distinct food arrival pant hoot?. *American Journal of Primatology*, 31(2), 99-109.
- Clay, Z., Archbold, J., & Zuberbühler, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ*, 3, e1124.
- Clements, M. (1998). Digital signal acquisition and representation. In S. L Hopp., M. J. Owren & C. S. Evans (Eds.), *Animal Acoustic Communication* (pp. 27-57). Springer.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Colonnesi, C., Zijlstra, B. J. H., van der Zande, A., & Bögels, S. M. (2012). Coordination of gaze, facial expressions and vocalizations of early infant communication with mother and father. *Infant Behavior and Development*, 35(3), 523–532.
- Corballis, M. C. (2017). *The evolution of language*. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbooks in psychology. APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior* (pp. 273–297). American Psychological Association
- Corballis, M. C. (2019). Mental time travel, language, and evolution. *Neuropsychologia*, 134, 107202.
- Crelin, E. S. (1987). *The human vocal tract: Anatomy, function, development, and evolution*. Vantage Press.
- Crockford, C. (2019). Why does the chimpanzee vocal repertoire remain poorly understood?-and what can be done about it. In C. Boesch & R. Witting (Eds.), *The chimpanzees of the Tai forest: 40 years of research* (pp. 394-409). Cambridge University Press.
- Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, 142(4) 397-421.

- Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science*, 5(5), 172066.
- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild Chimpanzees Produce Group-Specific Calls: A Case for Vocal Learning? *Ethology*, 110(3), 221–243.
- Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology*, 22(2), 142–146.
- Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization draws others' attention: A playback experiment with wild chimpanzees. *Animal Cognition*, 18(3), 581–591.
- Crockford, C., Wittig, R. M., & Zuberbühler, K. (2017). Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science Advances*, 3(11), e1701742.
- Davila-Ross, M., Allcock, B., Thomas, C., & Bard, K. A. (2011). Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion*, 11(5), 1013-1020.
- Davila-Ross, M., J Owren, M., & Zimmermann, E. (2009). Reconstructing the Evolution of Laughter in Great Apes and Humans. *Current Biology*, 19(13), 1106–1111.
- Dawkins, M. S. (2006). Through animal eyes: What behaviour tells us. *Applied Animal Behaviour Science*, 100(1), 4–10.
- Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain* (No. 202). WW Norton & Company.
- De Lathouwers, M., & Van Elsacker, L. (2006). Comparing infant and juvenile behavior in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*): a preliminary study. *Primates*, 47(4), 287-293.
- Delgado, C. E., Messinger, D. S., & Yale, M. E. (2002). Infant responses to direction of parental gaze: A comparison of two still-face conditions. *Infant Behavior and Development*, 25(3), 311-318.
- De Waal, F. B. (1988). The communicative repertoire of captive bonobos (*Pan paniscus*), compared to that of chimpanzees. *Behaviour*, 106(3-4), 183-251.
- Dezecache, G., Crockford, C., & Zuberbühler, K. (2019). The development of communication in alarm contexts in wild chimpanzees. *Behavioral Ecology and Sociobiology*, 73(8), 104.
- Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2020). A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. *Animal Cognition*, 1-13.
- Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2020). Flexibility in wild infant chimpanzee vocal behavior. *Journal of Language Evolution*, 1-17.
- Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science*, 23(1), e12843.

- D’Odorico, L., & Cassibba, R. (1995). Cross-sectional study of coordination between infants’ gaze and vocalizations towards their mothers. *Early Development and Parenting*, 4(1), 11–19.
- D’Odorico, L., Cassibba, R., & Salerni, N. (1997). Temporal Relationships Between Gaze and Vocal Behavior in Prelinguistic and Linguistic Communication. *Journal of Psycholinguistic Research*, 26(5), 539–556.
- Doyle, L. R., McCowan, B., Johnston, S., & Hanser, S. F. (2011). Information theory, animal communication, and the search for extraterrestrial intelligence. *Acta Astronautica*, 68(3), 406–417.
- Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), 32–57.
- Durkin, K. (1995). *Developmental social psychology: From infancy to old age*. Blackwell publishing.
- Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(4), e1493.
- Esling, J. H. (2012). The articulatory function of the larynx and the origins of speech. *Annual Meeting of the Berkeley Linguistics Society*, 38, 121-149.
- Fedurek, P., Machanda, Z. P., Schel, A. M., & Slocombe, K. E. (2013). Pant hoot chorusing and social bonds in male chimpanzees. *Animal Behaviour*, 86(1), 189–196.
- Fedurek, P., & Slocombe, K. (2011). Primate Vocal Communication: A Useful Tool for Understanding Human Speech and Language Evolution? *Human Biology*, 83, 153–173.
- Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, 6(1), 38226.
- Feldman, R. (2012). Oxytocin and social affiliation in humans. *Hormones and Behavior*, 61(3), 380–391.
- Ferdowsian, H. R., Durham, D. L., Kimwele, C., Kranendonk, G., Otali, E., Akugizibwe, T., Mulcahy, J. B., Ajarova, L., & Johnson, C. M. (2011). Signs of Mood and Anxiety Disorders in Chimpanzees. *PLOS ONE*, 6(6), e19855.
- Fernández-Carriba, S., Loeches, Á., Morcillo, A., & Hopkins, W. D. (2002). Asymmetry in facial expression of emotions by chimpanzees. *Neuropsychologia*, 40(9), 1523–1533.
- Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. *Animal Behaviour*, 55(4), 799–807.
- Fischer, J., Cheney, D. L., & Seyfarth, R. M. (2000). Development of infant baboons’ responses to graded bark variants. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1459), 2317–2321.
- Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2), 143-149.

- Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews*, 82, 22–31.
- Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2017). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour*, 134, 229–237.
- Fischer, J., Wheeler, B. C., & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology*, 25(21), R1028–R1029.
- Fitch, W. T. (2018). The biology and evolution of speech: a comparative analysis. *Annual review of linguistics*, 4, 255–279.
- Fitch, W. T. (2005). The evolution of language: a comparative review. *Biology and Philosophy*, 20(2–3), 193–203.
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences*, 4(7), 258–267.
- Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2(12), e1600723.
- Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Iyer, S. N., Lee, C.-C., Lambert, B., & Oller, D. K. (2014). Effects of Parental Interaction on Infant Vocalization Rate, Variability and Vocal Type. *Language Learning and Development*, 10(3), 279–296.
- Fogel, A., Toda, S., & Kawai, M. (1988). Mother-infant face-to-face interaction in Japan and the United States: A laboratory comparison using 3-month-old infants. *Developmental Psychology*, 24(3), 398–406.
- Freeberg, T. M., Dunbar, R. I. M., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1597), 1785–1801.
- Freeman, D. L., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The modern language journal*, 92(2), 200–213.
- Funkhouser, J. A., Mayhew, J. A., & Mulcahy, J. B. (2018). Social network and dominance hierarchy analyses at Chimpanzee Sanctuary Northwest. *PLoS ONE*, 13(2), e0191898.
- Gamba, M., Friard, O., Riondato, I., Righini, R., Colombo, C., Miaretsoa, L., ... & Giacoma, C. (2015). Comparative analysis of the vocal repertoire of Eulemur: a dynamic time warping approach. *International Journal of Primatology*, 36(5), 894–910.
- Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (Eds.). (1989). *Teaching sign language to chimpanzees*. Suny Press.
- Gautier, J. P. (1974). Field and Laboratory Studies of the Vocalizations of Talapoin Monkeys (*Miopithecus Talapoin*). *Behaviour*, 51(3–4), 209–271.
- Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PloS One*, 9(1), e84738.

- Gifford, I. I. I., Hauser, M. D., & Cohen, Y. E. (2003). Discrimination of Functionally Referential Calls by Laboratory-Housed Rhesus Macaques: Implications for Neuroethological Studies. *Brain, Behavior and Evolution*, 61(4), 213–224.
- Gómez-Robles, A., Hopkins, W. D., & Sherwood, C. C. (2014). Modular structure facilitates mosaic evolution of the brain in chimpanzees and humans. *Nature Communications*, 5(1), 1-9.
- Goodall, F. G. (1990). *The Chimpanzees of Gombe: Patterns of Behaviour*. Harvard University Press.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge University Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515-523.
- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, 80(3), 636-644.
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, 113(1), 52-58.
- Gómez, J. C. (1996). Ostensive behavior in great apes: The role of eye contact. In A. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 131-151). Cambridge University Press.
- Gould, S. J. (1977). *Ontogeny and phylogeny*. Harvard University Press.
- Gouzoules, H., & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, *Macaca nemestrina*, agonistic screams. *Animal Behaviour*, 37(3), 383–401.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388.
- Griebel, U., Pepperberg, I. M., & Oller, D. K. (2016). Developmental Plasticity and Language: A Comparative Perspective. *Topics in Cognitive Science*, 8(2), 435–445.
- Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal Responsiveness and the Development of Directed Vocalizing in Social Interactions. *Infancy*, 19(4), 385–408.
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B: Biological Sciences*, 275(1653), 2803–2811.
- Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, 21(5), 531–540.
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology*, 5(401), 1-12.

- Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of comparative and physiological psychology*, 45(5), 450.
- Henningsen, A., & Henningsen, M. A. (2020). Package 'censReg'.
- Hickey, R. (1984). Coronal segments in Irish English. *Journal of Linguistics*, 20(2), 233-250.
- Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia*, 46(9), 2423–2430.
- Hockett, C. F. (1990). A Comment on Design Features. *Anthropological Linguistics*, 32(3/4), 361–363.
- Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, 203(3), 88-97.
- Hopkins, W. D., & Savage-Rumbaugh, E. S. (1991). Vocal communication as a function of differential rearing experiences in Pan paniscus: A preliminary report. *International Journal of Primatology*, 12(6), 559-583.
- Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, 73(2), 281–286.
- Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2011). Do chimpanzees have voluntary control of their facial expressions and vocalizations. In A. Vilain, J. Schwartz, C, Abry, & J. Vauclair. (Eds.), *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans* (pp 71 - 88). John Benjamins Publishing Company.
- Hopp, S. L., Owren, M. J., & Evans, C. S. (Eds.). (2012). *Animal acoustic communication: Sound analysis and research methods*. Springer.
- Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, 3(2), 150–166.
- Hsu, H. C., & Fogel, A. (2001). Infant vocal development in a dynamic mother-infant communication system. *Infancy*, 2(1), 87-109.
- Hurford, J. R. (2003). The language mosaic and its evolution. In M. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 38-57). Oxford University Press.
- Itakura, S. (1993). Emotional Behavior during the Learning of a Contingency Task in a Chimpanzee. *Perceptual and Motor Skills*, 76(2), 563–566.
- Iyer, S. N., Denson, H., Lazar, N., & Oller, D. K. (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). *Clinical Linguistics & Phonetics*, 30(6), 470–488.
- Jhang, Y., & Oller, D. K. (2017). Emergence of Functional Flexibility in Infant Vocalizations of the First 3 Months. *Frontiers in Psychology*, 8(300), 1-11.
- Jones, R. (2013). *Communication in the real world: An introduction to communication studies*. The Saylor Foundation.
- Kano, F., Tanaka, M., & Tomonaga, M. (2008). Enhanced recognition of emotional stimuli in the chimpanzee (Pan troglodytes). *Animal cognition*, 11(3), 517-524.

- Kano, F., & Tomonaga, M. (2010). Attention to emotional scenes including whole-body expressions in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 124(3), 287-294.
- Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face orientation: two factors controlling apes' begging behavior from humans. *Animal cognition*, 7(4), 216-223.
- Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109(2), 224–234.
- Kershenbaum, A., Déaux, É. C., Habib, B., Mitchell, B., Palacios, V., Root-Gutteridge, H., & Waller, S. (2018). Measuring acoustic complexity in continuously varying signals: how complex is a wolf howl?. *Bioacoustics*, 27(3), 215-229.
- Keverne, E. B., Martensz, N. D., & Tuite, B. (1989). Beta-endorphin concentrations in cerebrospinal fluid of monkeys are influenced by grooming relationships. *Psychoneuroendocrinology*, 14(1-2), 155-161.
- Kirchhof, J., & Hammerschmidt, K. (2006). Functionally Referential Alarm Calls in Tamarins (*Saguinus fuscicollis* and *Saguinus mystax*)—Evidence from Playback Experiments. *Ethology*, 112(4), 346–354.
- Kobayashi, H., & Hashiya, K. (2011). The gaze that grooms: contribution of social factors to the evolution of primate eye morphology. *Evolution and Human Behavior*, 32(3), 157-165.
- Koda, H. (2004). Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (*Macaca fuscata yakui*). *Behaviour*, 141(10), 1279–1296.
- Kojima, S. (2008). Early vocal development in a chimpanzee infant. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 190-196). Springer.
- Kojima, S., Tatsumi, I. F., Kiritani, S., & Hirose, H. (1989). Vocal-auditory functions of the chimpanzee: Consonant perception. *Human Evolution*, 4(5), 403–416.
- Krause, M. A., & Beran, M. J. (2020). Words matter: Reflections on language projects with chimpanzees and their implications. *American Journal of Primatology*, 82(10), e23187.
- Krause, M. A., Udell, M. A., Leavens, D. A., & Skopos, L. (2018). Animal pointing: Changing trends and findings from 30 years of research. *Journal of Comparative Psychology*, 132(3), 326-345.
- Kretzschmar, W. A., & Kretzschmar Jr, W. A. (2015). *Language and complex systems*. Cambridge University Press.
- Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354(6308), 110–114.
- Krützen, M., Willems, E. P., & van Schaik, C. P. (2011). Culture and geographic variation in orangutan behavior. *Current Biology*, 21(21), 1808-1812.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160.

- Lameira, A. R. (2018). Origins of human consonants and vowels: articulatory continuities with great apes. In L. Boë, J. Fagot, P. Perrier, & J. Schwartz. (Eds.), *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates* (pp 75 – 100). Peter Lang Edition
- Lameira, A. R., & Call, J. (2018). Time-space–displaced responses in the orangutan vocal system. *Science Advances*, 4(11), eaau3401.
- Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H., ... & Wich, S. A. (2013). Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *The Journal of the Acoustical Society of America*, 134(3), 2326-2335.
- Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific reports*, 6(1), 1-10.
- Lameira, A. R., Hardus, M. E., Nouwen, K. J. J. M., Topelberg, E., Delgado, R. A., Spruijt, B. M., Sterck, E. H. M., Knott, C. D., & Wich, S. A. (2013). Population-Specific Use of the Same Tool-Assisted Alarm Call between Two Wild Orangutan Populations (*Pongopygmaeus wurmbii*) Indicates Functional Arbitrariness. *PLOS ONE*, 8(7), e69749.
- Langacker, R. W. (1976). A note on Uto-Aztecan consonant gradation. *International Journal of American Linguistics*, 42(4), 374-379.
- Laporte, M. N., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. *Developmental Science*, 14(5), 1220-1234.
- Lavelli, M., & Fogel, A. (2005). Developmental Changes in the Relationship Between the Infant’s Attention and Emotion During Early Face-to-Face Communication: The 2-Month Transition. *Developmental Psychology*, 41(1), 265–280.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, 1(1), 3.
- Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. *Animal Behavior and Cognition*, 7(2), 131-139.
- Lester, B. M., Boukydis, C. F. Z., Garcia-Coll, C. T., Peucker, M., McGrath, M. M., Vohr, B. R., Brem, F., & Oh, W. (1995). Developmental Outcome as a Function of the Goodness of Fit Between the Infant’s Cry Characteristics and the Mother’s Perception of Her Infant’s Cry. *Pediatrics*, 95(4), 516–521.
- Levréro, F., & Mathevon, N. (2013). Vocal Signature in Wild Infant Chimpanzees. *American Journal of Primatology*, 75(4), 324–332.
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
- Lieblich, A. K., Symmes, D., Newman, J. D., & Shapiro, M. (1980). Development of the Isolation Peep in laboratory-bred squirrel monkeys. *Animal Behaviour*, 28(1), 1–9.
- Liszkowski, U., Schäfer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science*, 20(5), 654-660.

- Locke, J. L. (2017). Emancipation of the voice: Vocal complexity as a fitness indicator. *Psychonomic Bulletin & Review*, 24(1), 232–237.
- Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, 29(3), 259-325.
- Lonsdorf, E. V., Anderson, K. E., Stanton, M. A., Shender, M., Heintz, M. R., Goodall, J., & Murray, C. M. (2014). Boys will be boys: Sex differences in wild infant chimpanzee social interactions. *Animal Behaviour*, 88, 79–83.
- Lyn, H., Russell, J. L., Leavens, D. A., Bard, K. A., Boysen, S. T., Schaeffer, J. A., & Hopkins, W. D. (2014). Apes communicate about absent and displaced objects: methodology matters. *Animal Cognition*, 17(1), 85-94.
- Maciej, P., Ndao, I., Hammerschmidt, K., & Fischer, J. (2013). Vocal communication in a complex multi-level society: Constrained acoustic structure and flexible call usage in Guinea baboons. *Frontiers in Zoology*, 10(1), 58.
- MacDonald, S. E., & Ritvo, S. (2016). Comparative cognition outside the laboratory. *Comparative Cognition & Behavior Reviews*, 11, 49-62
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281-297.
- Marler, P. (1967). Animal Communication Signals: We are beginning to understand how the structure of animal signals relates to the function they serve. *Science*, 157(3790), 769-774.
- Marler, P. (1976). Social organization, communication and graded signals: The chimpanzee and the gorilla. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology*. Cambridge University Press.
- Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees?. *Animal Behaviour*, 58(4), 825-830.
- Martins, E. P. (Ed.). (1996). *Phylogenies and the comparative method in animal behavior*. Oxford University Press.
- Matsusaka, T. (2004). When does play panting occur during social play in wild chimpanzees? *Primates*, 45(4), 221–229.
- Maynard-Smith, J. M., & Harper, D. (2003). *Animal Signals*. Oxford University Press.
- Maynard-Smith, J., & Harper, D. G. C. (1995). Animal Signals: Models and Terminology. *Journal of Theoretical Biology*, 177(3), 305–311.
- McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: Information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57(2), 409–419.
- Melis, A. P., Hare, B., & Tomasello, M. (2006). Engineering cooperation in chimpanzees: Tolerance constraints on cooperation. *Animal Behaviour*, 72(2), 275–286.
- Miles, H. L. W. (1990). *The cognitive foundations for reference in a signing orangutan*. In S. T. Parker & K. R. Gibson (Eds.), *"Language" and intelligence in monkeys and apes*:

- Comparative developmental perspectives* (p. 511–539). Cambridge University Press.
- Mitchell, P. R., & Kent, R. D. (1990). Phonetic variation in multisyllable babbling. *Journal of Child Language*, 17(2), 247–265.
- Mitani, J. C., & Brandt, K. L. (1994). Social Factors Influence the Acoustic Variability in the Long-distance Calls of Male Chimpanzees. *Ethology*, 96(3), 233–252.
- Mitani, J. C., Gros-Louis, J., & Macedonia, J. M. (1996). Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. *International Journal of Primatology*, 17(4), 569–583.
- Mitani, J. C., Hunley, K. L., & Murdoch, M. E. (1999). Geographic variation in the calls of wild chimpanzees: A reassessment. *American Journal of Primatology*, 47(2), 133–151.
- Mitani, J. C., Hasegawa, T., Gros-Louis, J., Marler, P., & Byrne, R. (1992). Dialects in wild chimpanzees? *American Journal of Primatology*, 27(4), 233–243.
- Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299–3303.
- Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 639–648.
- Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition*, 19(1), 223–231.
- Moore, J., Black, J., Hernandez-Aguilar, R. A., Idani, G. I., Piel, A., & Stewart, F. (2017). Chimpanzee vertebrate consumption: savanna and forest chimpanzees compared. *Journal of Human Evolution*, 112, 30–40.
- Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature reviews genetics*, 8(12), 943–949.
- Scherer, K. R., Mortillaro, M., & Mehu, M. (2013). Understanding the mechanisms underlying the production of facial expression of emotion: A componential perspective. *Emotion Review*, 5(1), 47–53.
- Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., ... & Kiliarntas, K. (2015). Embodied intersubjective engagement in mother–infant tactile communication: a cross-cultural study of Japanese and Scottish mother–infant behaviors during infant pick-up. *Frontiers in Psychology*, 6(66), 1–13.
- Newson, L. (2018). Cooperative Child Rearing: Evolution of Alloparenting in Hominins. *The International Encyclopedia of Anthropology*, 12, 1–7.
- Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2003). Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Sciences*, 100(12), 6930–6933.

- Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: Evolution of face flattening and speech. *Journal of Human Evolution*, 51(3), 244–254.
- Mason, N. A., Burns, K. J., Tobias, J. A., Claramunt, S., Seddon, N., & Derryberry, E. P. (2017). Song evolution, speciation, and vocal learning in passerine birds. *Evolution*, 71(3), 786-796.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.
- Oller, D. K. (2012). Underpinnings for a Theory of Communicative Evolution. In D. K. Oller & U. Griebel. (Eds.), *Evolution of Communication Systems: A Comparative Approach* (pp 49 – 67). MIT Press.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, 110(16), 6318-6323.
- Oller, D. K., & Griebel, U. (2014). On Quantitative Comparative Research in Communication and Language Evolution. *Biological Theory*, 9(3), 296–308.
- Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in psychology*, 10(729), 1-18.
- Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal development as a guide to modeling the evolution of language. *Topics in cognitive science*, 8(2), 382-392.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell’s Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, 4(11), e7808.
- Owings, D. H., & Morton, E. S. (1998). *Animal Vocal Communication: A New Approach*. Cambridge University Press.
- Owren, M. J., Rendall, D., & Ryan, M. J. (2010). Redefining animal signaling: influence versus information in communication. *Biology & Philosophy*, 25(5), 755-780.
- Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Duke University Press.
- Plooij, F. X. (1984). *The behavioral development of free-living chimpanzee babies and infants*. Ablex.
- Parr, L. A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees (Pan troglodytes). *Animal Cognition*, 4(3-4), 223-229.
- Parr, L. A., & Hopkins, W. D. (2000). Brain temperature asymmetries and emotional perception in chimpanzees, Pan troglodytes. *Physiology & Behavior*, 71(3), 363–371.
- Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion*, 7(1), 172–181.
- Patterson, F. G., & Cohn, R. H. (1990). Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development. *Word*, 41(2), 97-143.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, 1(2007), 224-239.

- Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95(2), 201–236.
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends in Cognitive Sciences*, 20(4), 304–318.
- Pollick, A. S., & de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, 104(19), 8184–8189.
- Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, 6(4), 486–494.
- Port, R. F., & Leary, A. P. (2005). Against Formal Phonology. *Language*, 81(4), 927–964.
- Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., Veeramah, K. R., Woerner, A. E., O'Connor, T. D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A. E., Malig, M., Hernandez-Rodriguez, J., ... Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, 499(7459), 471–475.
- Prat, Y., Taub, M., & Yovel, Y. (2015). Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances*, 1(2), e1500019.
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, 5(1), 1–11.
- Pusey, A. E. (1983). Mother-offspring relationships in chimpanzees after weaning. *Animal Behaviour*, 31(2), 363–377.
- Range, F., & Fischer, J. (2004). Vocal Repertoire of Sooty Mangabeys (*Cercocebus torquatus atys*) in the Taï National Park. *Ethology*, 110(4), 301–321.
- Raval, V., Goldberg, S., Atkinson, L., Benoit, D., Myhal, N., Poulton, L., & Zwiers, M. (2001). Maternal attachment, maternal responsiveness and infant attachment. *Infant Behavior and Development*, 24(3), 281–304.
- Ravignani, A., Fitch, W., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., ... & de Boer, B. (2016). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Frontiers in neuroscience*, 10, 274.
- Reddy, V. (2010). *How Infants Know Minds*. Harvard University Press.
- Reddy, V. (2019). Meeting infant affect. *Developmental psychology*, 55(9), 2020.
- Rendall, D., & Owren, M. J. (2013). Communication without meaning or information: abandoning language-based and informational constructs in animal communication theory. In U. E. Stegmann (Ed.), *Animal communication theory: information and influence* (pp. 151–188). Cambridge University Press.
- Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour*, 57(3), 583–592.

- Reynolds, V. (2005). *The Chimpanzees of the Budongo forest: Ecology, behaviour and conservation*. Oxford University Press.
- Richmond, V. P., & Martin, M. M. (1998). Socio-communicative style. In J. C. McCrosky (Ed.), *Communication and personality: Trait perspectives* (pp. 133-148). Hampton Press.
- Riede, T., Owren, M. J., & Arcadi, A. C. (2004). Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): frequency jumps, subharmonics, biphonation, and deterministic chaos. *American Journal of Primatology*, 64(3), 277-291.
- Robbins, M. M., Ando, C., Fawcett, K. A., Grueter, C. C., Hedwig, D., Iwata, Y., ... & Yamagiwa, J. (2016). Behavioral variation in gorillas: evidence of potential cultural traits. *PLoS One*, 11(9), e0160483.
- Romero, T., & de Waal, F. B. (2011). Third-party postconflict affiliation of aggressors in chimpanzees. *American Journal of Primatology*, 73(4), 397-404.
- Rosati, A. G., Hagberg, L., Enigk, D. K., Otali, E., Thompson, M. E., Muller, M. N., Wrangham, R. W., & Machanda, Z. P. (2020). Social selectivity in aging wild chimpanzees. *Science*, 370(6515), 473–476.
- Rosati, A. G., Wobber, V., Hughes, K., & Santos, L. R. (2014). Comparative developmental psychology: How is human cognitive development unique?. *Evolutionary Psychology*, 12(2), 448-473
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.
- Saussure, F. de. (1989). *Cours de linguistique générale*. Otto Harrassowitz Verlag.
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the society for research in child development*, 58(3/4), 1-252.
- Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2016). Call combinations, vocal exchanges and interparty movement in wild bonobos. *Animal Behaviour*, 122, 109–116.
- Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, 86(5), 955–965.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PLoS ONE*, 8(10).
- Schlenker, P., Chemla, E., Cäsar, C., Ryder, R., & Zuberbühler, K. (2017). Titi semantics: Context and meaning in Titi monkey call sequences. *Natural Language & Linguistic Theory*, 35(1), 271-298.
- Schoeninger, M. J., Moore, J. I. M., & Sept, J. M. (1999). Subsistence strategies of two “savanna” chimpanzee populations: The stable isotope evidence. *American Journal of Primatology*, 49(4), 297-314.

- Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2020). Was That a Scream? Listener Agreement and Major Distinguishing Acoustic Features. *Journal of Nonverbal Behavior*, 44(2), 233–252.
- Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, 18(12), 618–620.
- Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of Evolutionary Biology*, 21(2), 387–395.
- Scott-Phillips, T. (2015). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Macmillan International Higher Education.
- Seyfarth, R., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current opinion in behavioral sciences*, 21, 56-61
- Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, 34(6), 1640–1658.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210(4471), 801–803.
- Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, 80(1), 3–8.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.
- Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, 5(3), 225–279.
- Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology*, 119(1), 67-77.
- Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, 104(43), 17228–17233.
- Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. In E. Lonsdorf, S. Ross, & T. Matsuzawa (Eds.), *The Mind of the Chimpanzee: Ecological and Experimental Perspectives*, (pp. 192-207). Chicago, USA: University of Chicago Press.
- Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour*, 81(5), 919–924.
- Slocombe, K. E., & Zuberbühler, K. (2005). Functionally Referential Communication in a Chimpanzee. *Current Biology*, 15(19), 1779–1784.

- Snowdon, C. T. (1988). Communications as Social Interaction: Its Importance in Ontogeny and Adult Behavior. In D. Todt, P. Goedeeking, & D. Symmes (Eds.), *Primate Vocal Communication* (pp. 108–122). Springer.
- Snowdon, C. T., Cleveland, J., & French, J. A. (1983). Responses to context- and individual-specific cues in cotton-top tamarin long calls. *Animal Behaviour*, 31(1), 92–101.
- Snowdon, C. T. (2012). Social Processes in the Evolution of Complex Cognition and Communication. In D. K. Oller & U. Griebel(Eds.), *Evolution of Communication Systems: A comparative approach* (pp. 131-151). MIT Press.
- Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Mental Health Journal*, 2(2), 118–128.
- Suzuki, T. N., & Zuberbühler, K. (2019). Animal syntax. *Current Biology*, 29(14), 669-671.
- Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, 126(1), 354-366.
- Tadić, V., Pring, L., & Dale, N. (2010). Are language and social communication intact in children with congenital visual impairment at school age?. *Journal of Child Psychology and Psychiatry*, 51(6), 696-705.
- Taglialatela, J. P., Reamer, L., Schapiro, S. J., & Hopkins, W. D. (2012). Social learning of a communicative signal in captive chimpanzees. *Biology Letters*, 8(4), 498–501.
- Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., Schapiro, S. J., & Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American Journal of Primatology*, 77(11), 1143–1148.
- Taglialatela, J. P., Savage-Rumbaugh, S., & Baker, L. A. (2003). Vocal Production by a Language-Competent Pan paniscus. *International Journal of Primatology*, 24(1), 1–17.
- Tomasello, M., & Call, J. (2007). *Ape gestures and the origins of language*. In J. Call & M. Tomasello (Eds.), *The gestural communication of apes and monkeys* (p. 221–239). Taylor & Francis Group/Lawrence Erlbaum Associates.
- Tomasello, M., & Call, J. (2019). Thirty years of great ape gestures. *Animal cognition*, 22(4), 461-469.
- Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of communication*, 1(2), 223-259.
- Tomasello, M., & Gonzalez-Cabrera, I. (2017). The Role of Ontogeny in the Evolution of Human Cooperation. *Human Nature*, 28(3), 274–288.
- Townsend, S. W., & Manser, M. B. (2013). Functionally Referential Communication in Mammals: The Past, Present and the Future. *Ethology*, 119(1), 1–11.
- Townsend, S. W., & Zuberbuhler, K. (2009). Audience effects in chimpanzee copulation calls. *Communicative & Integrative Biology*, 2(3), 282–284.
- Ulving, T. (1953). Consonant Gradation in Eskimo. *International Journal of American Linguistics*, 19(1), 45–52.

- van de Rijt-Plooij, H. H., & Plooij, F. X. (1987). Growing independence, conflict and learning in mother-infant relations in free-ranging chimpanzees. *Behaviour*, 101(1-3), 1-86.
- van Hooff, J.A.R.A.M., 1973. A structure analysis of the social behaviour of a semi-captive group of chimpanzees. In M. von Cranach & I. Vine (Eds.), *Social Communication and Movement*, (pp.75-162). Academic Press.
- van Leeuwen, E. J. C., Cronin, K. A., & Haun, D. B. M. (2018). Population-specific social dynamics in chimpanzees. *Proceedings of the National Academy of Sciences*, 115(45), 11393–11400. <https://doi.org/10.1073/pnas.1722614115>
- Van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., ... & Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science*, 299(5603), 102-105.
- Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57–68.
- Vick, S. J., Waller, B. M., Parr, L. A., Smith Pasqualini, M. C., & Bard, K. A. (2007). A Cross-species Comparison of Facial Morphology and Movement in Humans and Chimpanzees Using the Facial Action Coding System (FACS). *Journal of Nonverbal Behavior*, 31(1), 1–20.
- Vihman, M. M. (2014). *Phonological Development: The First Two Years*. John Wiley & Sons.
- Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, 61(2), 397-445.
- Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires—Hard vs. soft classification approaches. *PloS One*, 10(4), e0125785.
- Waller, B. M., Whitehouse, J., & Micheletta, J. (2017). Rethinking primate facial expression: A predictive framework. *Neuroscience & Biobehavioral Reviews*, 82, 13–21.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015). Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology*, 25(4), 495–499.
- Watts, D. P. (2016). Production of grooming-associated sounds by chimpanzees (Pan troglodytes) at Ngogo: variation, social learning, and possible functions. *Primates*, 57(1), 61-72.
- Watumull, J., Hauser, M. D., Roberts, I. G., & Hornstein, N. (2014). On recursion. *Frontiers in Psychology*, 8(1017), 1-7.
- Werner, E. E., & Gilliam, J. F. (1984). The Ontogenetic Niche and Species Interactions in Size-Structured Populations. *Annual Review of Ecology and Systematics*, 15, 393–425.

- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences*, 102(suppl 1), 6543–6549.
- Wharton, T. (2003). Natural Pragmatics and Natural Codes. *Mind & Language*, 18(5), 447–477.
- Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., ... & van Schaik, C. P. (2012). Call cultures in orang-utans?. *PLoS one*, 7(5), e36180.
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, 50(1), 56–64.
- Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., & Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*. *Animal Behaviour*, 123, 305–316.
- Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge University Press.
- Wonke, G., & Wallschläger, D. (2009). Song dialects in the yellowhammer *Emberiza citrinella*: bioacoustic variation between and within dialects. *Journal of Ornithology*, 150(1), 117–126.
- Yamanashi, Y. (2010). Emotional consequences when chimpanzees (*Pan troglodytes*) face challenges: Individual differences in self-directed behaviours during cognitive tasks. *Animal Welfare*, 19, 25–30
- Yoo, H., Bowman, D. A., & Oller, D. K. (2018). The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Frontiers in Psychology*, 9(1510), 1–15.
- Zec, D. (2005). Prosodic differences among function words. *Phonology*, 22(1), 77–112.
- Zadah, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
- Zadeh, L. A. (2008). Is there a need for fuzzy logic?. *Information sciences*, 178(13), 2751–2779
- Zeveloff, S. I., & Boyce, M. S. (1982). Why Human Neonates Are So Altricial. *The American Naturalist*, 120(4), 537–542.
- Zhou, K., Fu, C., & Yang, S. (2014). Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation. *Science China Information Sciences*, 57(11), 1–8.
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63(2), 293–299.
- Zuberbühler, K. (2011). Cooperative breeding and the evolution of vocal flexibility. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 71–83). Oxford University Press.
- Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37(2), 125–144
- Zwart, J. W. (2011). Recursion in Language: A Layered-Derivation Approach. *Biolinguistics*, 5(1–2), 043–056.

Appendices

Appendix 1 – Supplementary methods and results (Chapter 2)

Supplementary methods

Table S1. Number of calls per subject

Subject	Stage	Call type								Total
		Bark	Grunt	Hoo	Laugh	Pant	Scream	Squeak	Whimper	
						hoot				
Bill	Infant	0	43	3	0	0	10	5	31	81
Joey	Infant	1	17	0	0	0	0	0	4	22
Rogers	Infant	0	21	0	7	0	0	0	3	31
Murial	Infant	0	3	0	0	0	4	0	4	11
Victoria	Infant	0	45	0	2	0	0	0	4	51
Nancy	Infant	0	25	0	6	0	0	0	11	42
Dennis	Infant	1	4	0	3	0	2	0	3	13
Don	Infant	0	23	1	13	0	1	1	4	43
Mumba	Infant	0	2	1	0	0	1	0	5	9
Gloria	Infant	0	10	0	11	0	0	0	0	21
Ida	Infant	0	2	14	0	0	0	0	18	34
Tom	Infant	2	12	0	4	0	2	0	3	23
Tina	Infant	0	17	0	6	0	0	0	0	23
Brent	Juvenile	0	1	0	0	0	0	0	0	1
Chitalu	Juvenile	1	1	0	0	0	0	0	2	4
Jewel	Juvenile	0	31	1	20	0	0	1	0	53
May	Juvenile	0	6	0	7	0	2	0	5	20

Appendices

Danny	Juvenile	0	14	11	8	0	4	0	17	54
Little Jack	Juvenile	0	7	1	11	0	4	0	9	26
Martin	Juvenile	1	14	2	1	0	8	0	7	33
Ireen	Juvenile	0	2	0	1	0	0	0	0	3
Jones	Juvenile	1	22	1	12	2	5	0	2	45
Kenny	Juvenile	0	21	1	23	1	1	0	4	51
Lods	Juvenile	0	1	0	0	0	3	0	4	8
Bruce	Juvenile	1	16	4	1	1	1	0	1	25
Jack	Juvenile	0	9	0	3	0	0	0	0	12

Supplementary results

Supplementary Results

This section covers cluster stability, cluster composition, cluster typicality, and descriptive statistics on acoustic parameters. Cluster stability includes all stability measurements not reported in the main text **(a)**. Cluster composition includes cluster composition for the infant and juvenile counterpart models **(b)** and evidence that the coded calls were acoustically distinct **(c)**. Cluster typicality **(d-g)** includes a breakdown of typicality per call type and cluster for all models reported in the main text. The acoustic parameter overview includes descriptive statistics on all call types for infants and juveniles **(h)** and on the acoustic characteristics of call clusters in best fitting and counterpart models **(i)**.

a) Cluster stability measurements for all infant and juvenile models

For infants, stability measurements could not be calculated for 4-7 cluster solutions, suggesting cluster differentiations for these models were heavily reliant on a small

number of acoustic parameters. In contrast, stability measurements could be calculated for both 2 and 3 cluster solutions. Of the four different stability measurements taken, APN and ADM identified 2-cluster solutions as the most stable, whereas AD and FOM identified 3-cluster solutions as the most stable. The APN values are of particular note here, as these show that 2 and 3-cluster solutions showed a mean of only 2% and 6% of datapoints changed cluster when the models were recalculated with a single variable from the full dataset missing (table S1). This suggests that these cluster solutions did not rely on a small number of acoustic parameters.

Table S1. Mean stability measurements for all clusters for infants and juveniles.

	<u>K</u>					
Stability	2	3	4	5	6	7
measurements						
<hr/>						
Mean proportion						
of non-overlap						
(APN)						
Infant	0.024	0.061	N/A	N/A	N/A	N/A
Juvenile	0.045	N/A	N/A	N/A	N/A	N/A
<hr/>						
Mean distance						
(AD)						
Infant	3.519	3.190	N/A	N/A	N/A	N/A
Juvenile	3.892	N/A	N/A	N/A	N/A	N/A
<hr/>						

Mean distance						
between means						
(ADM)						
Infant	0.029	0.053	N/A	N/A	N/A	N/A
Juvenile	0.048	N/A	N/A	N/A	N/A	N/A
Figure of merit						
(FOM)						
Infant	0.781	0.772	N/A	N/A	N/A	N/A
Juvenile	0.874	N/A	N/A	N/A	N/A	N/A
<i>Note: Bold values indicate strongest evidence for cluster stability</i>						

For juveniles, 2-cluster solutions showed high levels of stability across all measurements and stability could not be calculated for all other solutions, suggesting solutions with > 2 clusters were dependent on a small number of acoustic parameters. In comparison to the stability of infant 2-cluster solutions, the stability of juvenile 2-cluster solutions was slightly lower, although cluster stability was still very high. For example, AVPN measurements showed that on mean 4.5% of datapoints changed their main cluster membership when the solution was systematically recalculated with 1 variable missing, compared to 2% for the infant and overall models (table S1).

b) Cluster compositions for the counterpart (non-best fitting) infant ($K=2$, $\mu=1.1$) and juvenile models ($K=2$, $\mu=1.5$)

Cluster compositions for the counterpart infant and juvenile models were highly similar to cluster compositions for the best fitting models. The only notable exception is for juvenile grunts, wherein the majority (52.95%) are now found in cluster 2 whereas the majority (69.33%) are found in cluster 1 for the best-fitting model.

Table S2 Total number of calls per call type per cluster and percentage of each call type per cluster for counterpart infant and juvenile models.

Cluster	<u>Call Types</u>						
	Grunt	Whimper	Laugh	Scream	Hoo	Bark	Squeak
Number							
1							
Infants	106	13	31	11	1	4	2
	(54.63%)	(8.28%)	(93.93%)	(100%)	(7.69%)	(100%)	(50%)
Juveniles	48	5	39	17	0	1	0
	(47.05%)	(8.33%)	(61.90%)	(89.47%)	(0%)	(100%)	(0%)
2							
Infants	88	144	2	0	12	0	2
	(45.37%)	(91.72%)	(6.07%)	(0%)	(92.31%)	(0%)	(50%)
Juveniles	54	56	24	2	16	0	2
	(52.95%)	(91.67%)	(38.10%)	(10.53%)	(100%)	(0%)	(100%)

c) Cluster compositions for an overall model $K=7$, $\mu=1.1$

To examine the extent to which the coded call types were separable from one another, cluster compositions were examined for the model $K=7$, $\mu=1.1$. $K=7$ was chosen because this corresponded to the number of call types coded and $\mu=1.1$ to allow minimal overlap between clusters. Table S3 shows that the call types that occupy the majority of the dataset (i.e. grunts, whimpers, laughs and screams) have mutually exclusive primary cluster memberships, indicating the coded call types are distinct from one another. The exception is laughs which shared a primary cluster membership with grunts. However, the second highest cluster memberships for laughs was only 4% lower and was not shared by any other call type. As such, the data nonetheless suggest laughs are largely distinct from the other coded call types.

Table S3. Cluster composition for the model $K=7$, $\mu=1.1$ as indicated by percentage of calls per call type per cluster.

Cluster	<u>Call Type</u>						
	Grunt	Whimper	Laugh	Scream	Hoo	Bark	Squeak
1	17.2	11.00	4.81	46.42	13	60.00	25.00
2	15.38	2.94	32.5	46.40	N/A	20.00	25.00
3	17.00	20.6	14.00	3.60	46.66	20.00	N/A
4	16.29	54.90	1.20	N/A	40	N/A	50
5	25.33	2.94	36.14	N/A	N/A	N/A	N/A
6	8.59	0.98	7.22	N/A	N/A	N/A	N/A
7	0.45	6.86	3.61	3.57	N/A	N/A	N/A

Note: The highest values per call type indicate primary cluster membership and are highlighted in bold.

d) Cluster typicality for best fitting infant model ($K=2$, $\mu=1.5$) and juvenile counterpart model with the same parameters

Since μ values constrain the overlap between clusters and thereby constrain typicality values, models were created using the same parameters for the purpose of direct infant-juvenile comparisons. Consistent with the observation that infant calls were less graded than juvenile calls, figure S1 below shows a clear right-skew in the distribution of typicality coefficients for infants. By contrast, juveniles appear to have a flatter distribution, consistent with increased acoustic gradedness of juvenile calls compared to infants. Right-skewed distribution of typicality measurements indicates less acoustic gradation because right-skewness indicates most of the typicality values were high, meaning low levels of overlap between clusters.

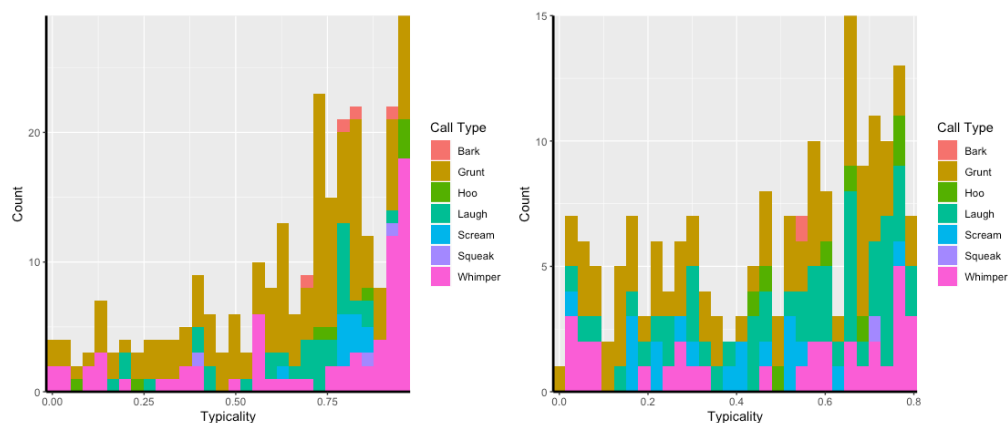


Figure S1. Histogram depicting the distribution of observed typicality coefficients and the number of calls that were observed in that range for the best fitting infant model (left) and the corresponding juvenile model (right) with the parameters $K=2$, $\mu=1.5$.

The difference between infant and juvenile call typicality was also consistent across all call types (table S4), with all mean typicality coefficients for infants being higher than mean typicality coefficients for juveniles. However, there was considerable variation in the difference between infant and juvenile mean typicality coefficient per call type, ranging between 0.01 – 0.38. Notably, for the cases wherein there was only a 0.01 difference in call typicality (squeaks), only a single call type was observed for juveniles.

Table S4. Mean and standard deviation of typicality overall and per call type and percentage of calls that met the typicality criteria for the best fitting infant model and the corresponding juvenile model.

Call Type	Mean (SD) Typicality	% Atypical calls	% Typical calls
Bark			
Infant	0.81 (0.10)	0	100
Juvenile	0.55 (N/A)	0	100
Grunt			
Infant	0.62 (0.24)	15.97	84.02
Juvenile	0.44(0.24)	28	72
Hoo			
Infant	0.68 (0.34)	25	75
Juvenile	0.61 (0.13)	0	100
Laugh			
Infant	0.63 (0.21)	12.12	87.87
Juvenile	0.51 (0.22)	15.68	84.31

Scream			
Infant	0.79 (0.06)	0	100
Juvenile	0.37 (0.19)	33.33	66.66
Squeak			
Infant	0.72 (0.28)	0	100
Juvenile	0.71 (N/A)	0	100
Whimper			
Infant	0.69 (0.31)	15.71	84.28
Juvenile	0.47 (0.28)	24.24	75.75
Total			
Infant	0.65 (0.26)	14.70	85.29
Juvenile	0.47 (0.24)	22.99	77.00

Note: N/A = No value could be calculated because there was either no call of that type in the cluster or only 1 call of that type in the cluster in the case of standard deviations.

e) Cluster typicality per cluster for best fitting infant model (K=2, =1.5) and the juvenile counterpart model with the same parameters

Since μ values constrain the overlap between clusters and thereby constrain typicality values, models were created using the same parameters for the purpose of direct infant-juvenile comparisons. For infants, a wide range of typicality coefficients was observed for each cluster, with cluster 1 ranging between 0.002-0.865, and cluster 2 ranging between 0.004-0.961 (figure S2). This shows that primary cluster membership was between 0.02% - 86.5% higher than secondary cluster membership for cluster 1 and between 0.4% and 96.1%

for cluster 2. For juveniles, both clusters showed similar ranges of typicality, with cluster 1 ranging from 0.013-0.79, and cluster 2 ranging between 0.015-0.807 (figure S2), indicating primary cluster membership was between 1.3% and 79% higher than secondary for cluster 1 and 1.5% and 80% for cluster 2.

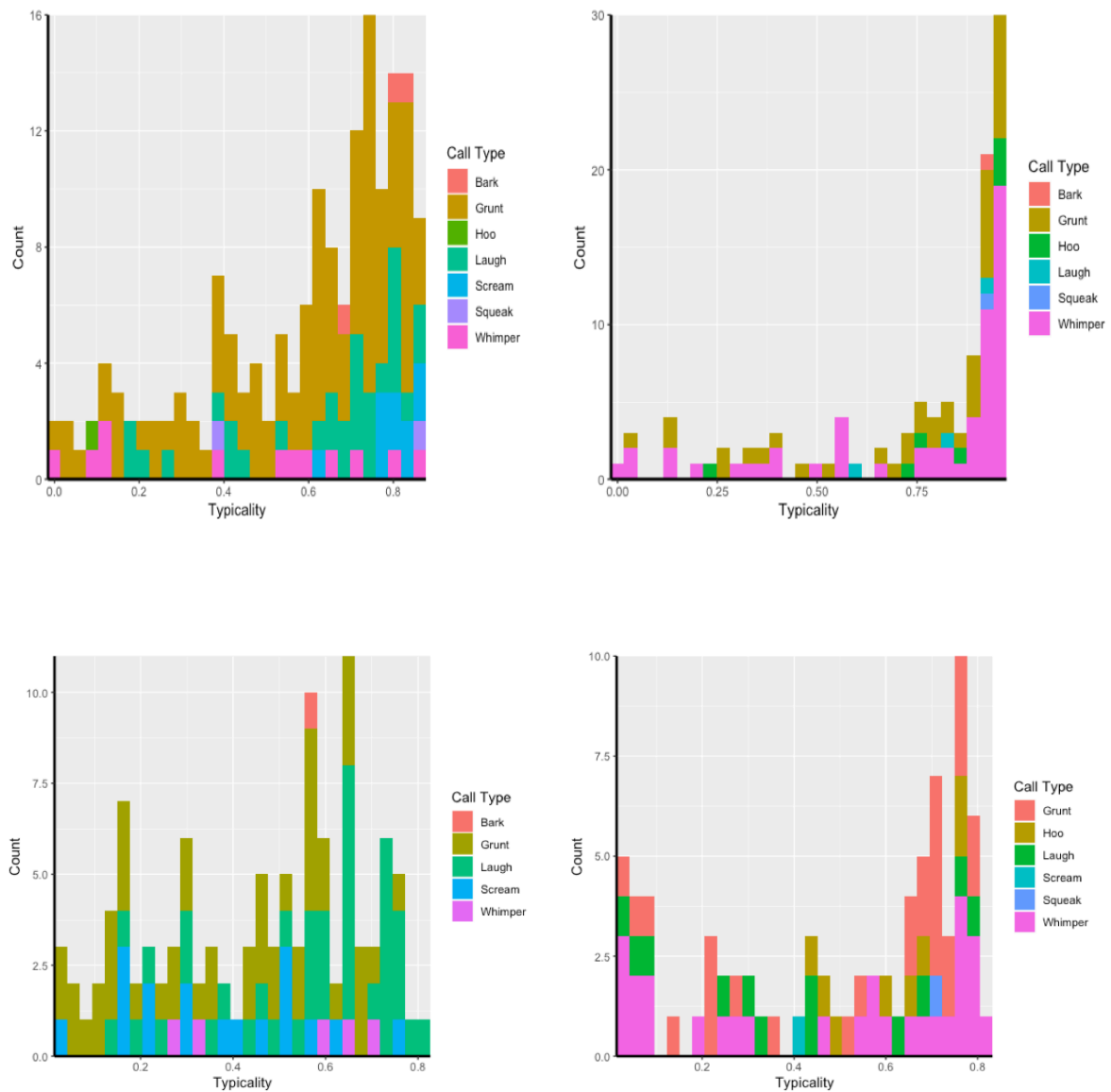


Figure S2. Histogram depicting the range of observed typicality coefficients and the number of calls that were observed in that range for cluster 1 infant model (top left), cluster 2 infant model (top right), cluster 1 juvenile model (bottom left) and cluster 2 juvenile model

(bottom right) for the best fitting infant model and the corresponding juvenile model. Bars are colour coded according to call type, showing the typicality of specific call types.

For infants, both clusters generated relatively high mean typicality coefficients (table S5). Cluster 2 generated higher mean typicality than cluster 1. The majority of calls were classified as typical and this pattern was consistent across call types with the exception of hoo calls in cluster 1. However, this was a single case. The percentage of calls classified as typical showed little difference between clusters. For juveniles, between cluster 1 and 2 the mean and standard deviation of cluster typicality was very similar but consistently lower than for infants (table S5). Similar to the infant model, the majority of calls were classified as typical in the juvenile model. Together, these findings suggest that while both cluster 1 and 2 are positively skewed towards typicality (cluster 2 in particular), this skew is weaker than in the corresponding infant model. Right-skewed distribution of typicality measurements indicates less acoustic gradation because right-skewness indicates most of the values high, meaning low levels of overlap between clusters.

Table S5. Mean and standard deviation of typicality for each cluster and per call type for the best fitting infant model and the corresponding juvenile model. Percentages of calls that met the typicality criteria are also shown.

Call Type	Cluster 1			Cluster 2		
	Mean (SD)	% Atypical	% Typical	Mean (SD)	% Atypical	% Typical
	Typicality			Typicality		
Bark						

Appendices

Infant	0.77 (0.07)	0	100	0.93 (N/A)	0	100
Juvenile	0.55 (N/A)	0	100	N/A	N/A	N/A
Grunt						
Infant	0.58 (0.22)	13.46	86.53	0.72 (0.27)	17.5	82.5
Juvenile	0.39 (0.22)	29.16	70.83	0.52 (0.26)	25.95	74.07
Hoo						
Infant	0.07 (N/A)	100	0	0.77 (0.25)	14.28	85.71
Juvenile	N/A	N/A	N/A	0.60 (0.13)	0	100
Laugh						
Infant	0.62 (0.21)	13.33	86.66	0.77 (0.17)	0	100
Juvenile	0.55 (0.19)	10.25	89.74	0.39 (0.27)	33.33	66.66
Scream						
Infant	0.79 (0.06)	0	100	N/A	N/A	N/A
Juvenile	0.36 (0.19)	23.52	76.47	0.40 (N/A)	0	100
Squeak						
Infant	0.62 (0.32)	0	100	0.92 (N/A)	0	100
Juvenile	N/A	N/A	N/A	0.71 (N/A)	0	100
Whimper						
Infant	0.45 (0.29)	33.33	66.66	0.74 (0.29)	13.79	86.20
Juvenile	0.51 (1.98)	0	100	0.47 (0.29)	28.57	71.42
Total						
Infant	0.59 (0.23)	14.19	85.80	0.74 (0.27)	14.54	85.45
Juvenile	0.45 (0.22)	20	80	0.49 (0.27)	24.67	75.32

Note: N/A = No value could be calculated because there was either no calls of that type in the cluster or only 1 call of that type in the cluster in the case of standard deviations.

f) Cluster typicality for best fitting juvenile model ($K=2$, $\mu=1.1$) and the infant counterpart model with the same parameters

Since μ values constrain the overlap between clusters and thereby constrain typicality values, models were created using the same parameters for the purpose of direct infant-juvenile comparisons. Consistent with the observation that there was no significant difference in typicality between infants and juveniles for the best fitting juvenile model and its infant counterpart, there is a clear strong right-skew in typicality coefficients for both models (figure S3). This is however, unsurprising since in the model $\mu=1.1$, meaning almost no overlap is permitted between clusters and therefore typicality coefficients (which reflect cluster overlap) are highly constrained.

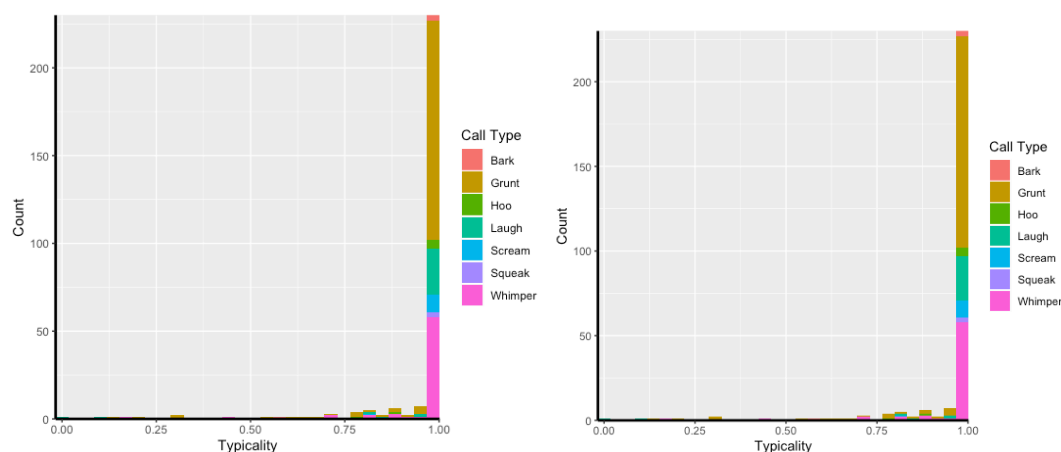


Figure S3. Histogram depicting the range of observed typicality coefficients and the number of calls that were observed in that range for the best-fitting juvenile model (left) and the corresponding infant model (right) with the parameters ($K=2$, $\mu=1.1$).

There was no consistent pattern in mean typicality across call types with respect to infants vs juveniles in these models (table S6). However, it is important to note that the range of distances between mean typicality coefficients was very small in this model relative to the best fitting infant model and its juvenile counterpart (0 - 0.13). Again, since $\mu=1.1$ in this model, almost no cluster overlap is permitted and it is therefore unsurprising that no clear differences between infants and juveniles were observed in these models.

Table S6. Mean and standard deviation of typicality overall and per call type and percentage of calls that met the typicality criteria for the best fitting juvenile model and the corresponding infant model.

Call Type	Mean (SD) Typicality	% Atypical calls	% Typical calls
Bark			
Infant	0.89 (0.21)	0	100
Juvenile	0.99 (N/A)	0	100
Grunt			
Infant	0.95 (0.14)	2.73	97.26
Juvenile	0.91 (0.17)	5.33	94.66
Hoo			
Infant	0.95 (0.08)	0	100
Juvenile	0.98 (0.01)	0	100
Laugh			
Infant	0.92 (0.23)	6.25	93.75

Juvenile	0.93 (0.17)	5.88	94.11
Scream			
Infant	0.98 (0.06)	0	100
Juvenile	0.97 (0.03)	0	100
Squeak			
Infant	0.99 (0.002)	0	100
Juvenile	0.99 (N/A)	0	100
Whimper			
Infant	0.95 (0.13)	2.89	97.10
Juvenile	0.82 (0.29)	18.18	81.81
Total			
Infant	0.95 (0.14)	2.94	97.05
Juvenile	0.91 (0.19)	6.95	93.04

Note: N/A = No value could be calculated because there was either no call of that type in the cluster or only 1 call of that type in the cluster in the case of standard deviations.

g) Cluster typicality per cluster for best fitting juvenile model ($K=2$, $\mu=1.1$) and the infant counterpart model with the same parameters

Since μ values constrain the overlap between clusters and thereby constrain typicality values, models were created using the same parameters for the purpose of direct infant-juvenile comparisons. For infants, within cluster 1 of the $K=2$, $\mu=1.1$ infant model, observed typicality coefficients ranged from 0.01-0.99, whereas cluster 2 ranged between 0.10-1

(figure S4). This indicates that while both clusters contained calls that were maximally different from the alternative cluster, the lowest observed coefficients show that the least typical calls had primary cluster memberships that were 9% higher for cluster 2 compared to cluster 1. For juveniles, within cluster 1 of the best fitting juvenile model, typicality coefficients ranged between 0.28-0.99, while within cluster 2 typicality ranged between 0.15-0.99 (figure S4), meaning primary cluster membership varied between 28% - 99% and 15% - 99% higher than secondary cluster membership for clusters 1 and 2 respectively.

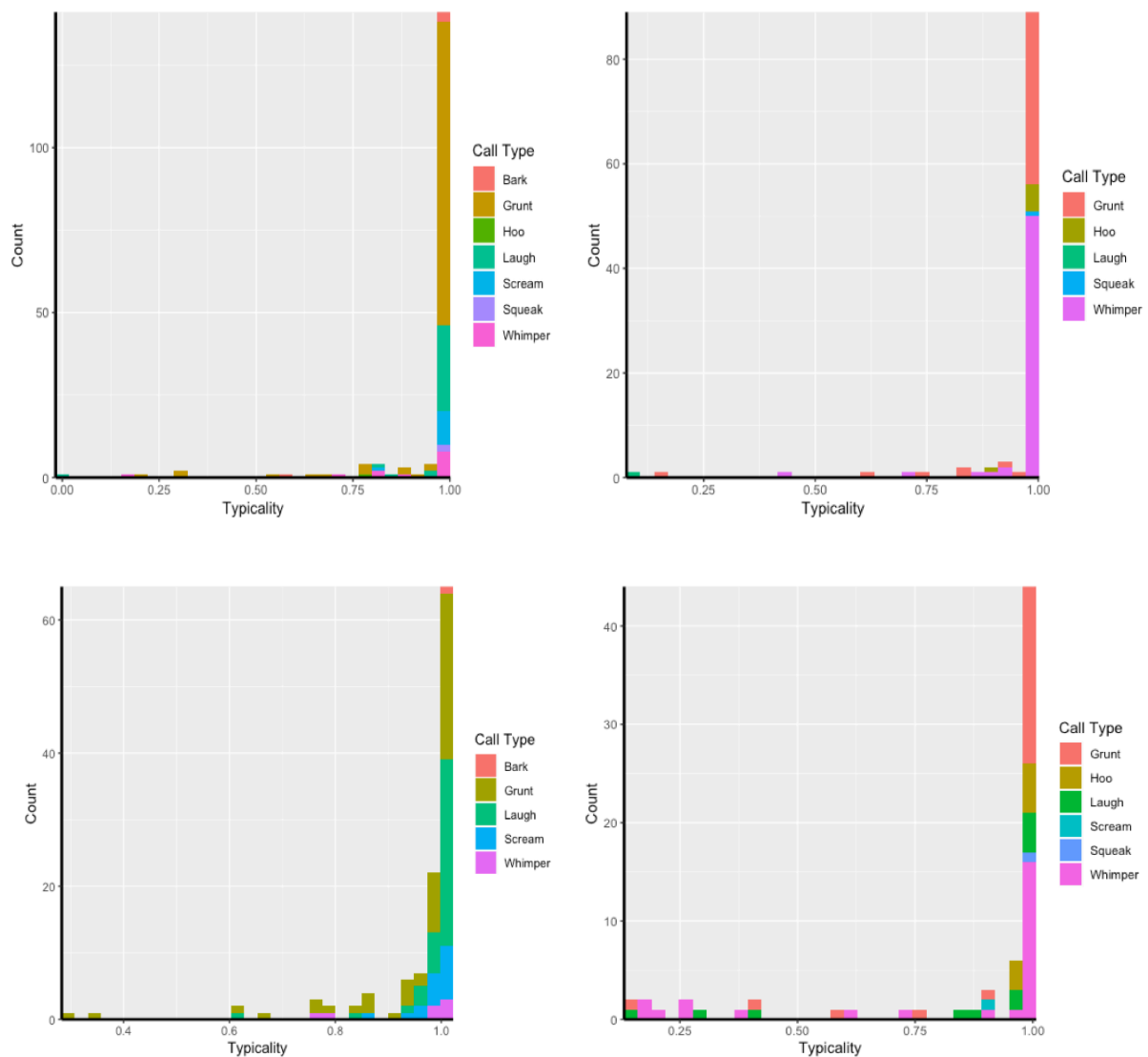


Figure S4. Histogram depicting the range of observed typicality coefficients and the number of calls that were observed in that range for cluster 1 infant model (top left), cluster 2 infant model (top right), cluster 1 juvenile model (bottom left) and cluster 2 juvenile model (bottom right) for the best fitting juvenile model and the corresponding infant model. Bars are colour coded according to call type, showing the typicality of specific call types.

For infants, cluster 1 and cluster 2 both generated high mean typicality coefficients which were almost equal (table S7) and the percentage of calls classified as typical for each cluster was also almost equal. Squeaks were the most typical call class in both clusters, indicating highly distinct squeak variants. However, only one squeak was observed in cluster 2 for infants and juveniles, suggesting that while squeaks appear to show considerable variability, the majority of squeaks had acoustic qualities that overlapped with the acoustic qualities of other calls that belonged to cluster 1. In both clusters, the majority of calls were classified as typical, and this finding was consistent across call types with the exception of laughs in cluster 2, which were 100% atypical for infants, although there was only one case of laughter in this cluster. For juveniles, both clusters showed high mean typicality, although cluster 1 showed 10% higher mean typicality than cluster 2 (table S7). The majority of calls were classified as typical, and this pattern was consistent across all call types, with the exception of hoo calls and squeaks in cluster 1 and barks in cluster 2 wherein no data points were available. Between juvenile clusters, cluster 1 had 14% more calls classified as typical than cluster 2.

Table S7. Mean and standard deviation of typicality for each cluster and per call type for the best fitting juvenile model and the corresponding infant model. Percentages of calls that met the typicality criteria are also shown.

Call Type	<u>Cluster 1</u>			<u>Cluster 2</u>		
	Mean (SD)	% Atypical	% Typical	Mean (SD)	%	% Typical
	Typicality			Typicality	Atypical	
Bark						
Infant	0.89 (0.21)	0	100	N/A	N/A	N/A
Juvenile	0.99 (N/A)	0	100	N/A	N/A	N/A
Grunt						
Infant	0.95 (0.13)	2.83	97.16	0.95 (0.15)	2.5	97.5
Juvenile	0.92 (0.15)	3.84	96.15	0.90 (0.22)	8.69	91.30
Hoo						
Infant	0.77 (N/A)	0	100	0.98 (0.04)	0	100
Juvenile	N/A	N/A	N/A	0.98 (0.01)	0	100
Laugh						
Infant	0.95 (0.17)	3.22	96.77	0.10 (N/A)	100	0
Juvenile	0.98 (0.06)	0	100	0.77 (0.32)	27.27	72.72
Scream						
Infant	0.98 (0.06)	0	100	N/A	N/A	N/A
Juvenile	0.98 (0.03)	0	100	0.91 (N/A)	0	100
Squeak						
Infant	0.99 (0.003)	0	100	0.99 (N/A)	0	100

Juvenile	N/A	N/A	N/A	0.99 (N/A)	0	100
Whimper						
Infant	0.87 (0.23)	7.69	92.30	0.97 (0.08)	1.78	98.21
Juvenile	0.93 (0.10)	0	100	0.74 (0.32)	23.07	76.92
Total						
Infant	0.94 (0.15)	2.97	97.02	0.95 (0.14)	2.88	97.11
Juvenile	0.95 (0.11)	1.70	98.29	0.85 (0.27)	15.71	84.28

Note: N/A = No value could be calculated because there was either no call of that type in the cluster or only 1 call of that type in the cluster in the case of standard deviation

h) Means and standard deviations of acoustics parameters per call type for infants and juveniles

	Acoustic Parameters													
Call Type	Low Freq (Hz)	Delta Freq (Hz)	Q1 Freq (Hz)	Q3 Freq (Hz)	Mean Entropy	BW 90% (Hz)	Center Freq (Hz)	Peak Freq (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Duration (s)	Element number	Element rate (s ⁻¹)
Grunt														
Infant	541.31 (1005.35)	9337.93 (5609.46)	1248.58 (1388.89)	2558.22 (2140.23)	3.33 (0.99)	3880.74 (2503.341)	1856.22 (2026.80)	1571.94 (1809.17)	1755.53 (1026.22)	3812.40 (2074.21)	5936.34 (3202.16)	1.16 (1.66)	2.28 (2.21)	11.63 (19.37)
Juvenile	306.74 (746.76)	10184.38 (5255.69)	878.03 (839.13)	2402.01 (1391.82)	3.53 (0.91)	4281.68 (2640.66)	1417.92 (1012.57)	1123.61 (1025.71)	1922.25 (1035.54)	4071.23 (1899.51)	6267.51 (2891.80)	1.62 (3.51)	3.07 (5.23)	8.75 (7.47)
Whimper														
Infant	140.31 (110.90)	3925.19 (4703.66)	431.74 (318.06)	916.12 (901.35)	1.95 (0.73)	1230.42 (1519.22)	630.90 (499.50)	524.52 (404.01)	614.93 (619.67)	1340.18 (1572.16)	2116.73 (2571.75)	22.25 (158.53)	8.14 (9.35)	3.80 (2.66)
Juvenile	256.65 (266.20)	8779.37 (5230.12)	830.41 (412.44)	1511.11 (728.22)	2.73 (0.65)	2378.23 (1568.81)	1096.05 (534.13)	1833.34 (4221.95)	1391.89 (786.81)	3091.53 (1796.19)	4850.01 (2813.68)	14.81 (25.69)	16.8 (19.41)	4.53 (9.49)
Laugh														
Infant	594.55 (1350.01)	11454.31 (339.36)	1653.76 (1417.10)	3978.07 (1678.51)	4.27 (0.33)	5840.86 (1309.98)	2587.19 (1504.48)	2139.25 (1489.94)	2207.61 (499.48)	4532.17 (995.33)	7024.11 (1532.20)	3.41 (3.57)	7.10 (5.84)	4.04 (1.95)
Juvenile	261.48 (449.12)	10372 (4306.75)	1127.92 (651.73)	3410.71 (1195.49)	4.12 (0.49)	5792.01 (1953.81)	1962.69 (822.71)	1762.92 (1384.44)	2078.63 (800.12)	4237.87 (1667.63)	6441.06 (2484.39)	3.60 (7.55)	8.98 (17.20)	7.72 (21.38)
Scream														
Infant	834.62 (156.85)	15432.15 (1815.68)	1833.33 (406.25)	3989.57 (821.46)	3.34 (0.37)	4937.5 (1278.57)	2687.5 (419.26)	1947.92 (755.51)	2795.92 (227.76)	5427.64 (359.82)	8215.55 (395.25)	2.14 (2.74)	5.1 (6.31)	3.83 (2.88)
Juvenile	585.55 (237.16)	15522.65 (2043.12)	1518.75 (439.29)	2480.38 (591.06)	2.88 (0.45)	2865.79 (1292.37)	1891.84 (604.50)	1845.13 (699.68)	2215.29 (437.32)	5009.44 (1066.61)	8068.44 (1690.01)	5.45 (6.45)	11.04 (11.77)	2.32 (0.92)
Hoo														
Infant	204.43 (127.10)	1143.96 (868.07)	312.5 (108.25)	562.5 (187.5)	1.65 (0.70)	812.5 (780.62)	500.00 (108.25)	500.00 (108.25)	448.04 (81.81)	777.69 (321.28)	1182.99 (468.81)	2.38 (3.76)	1.5 (0.57)	5.25 (3.50)
Juvenile	256.62 (299.19)	7882.07 (4014.20)	631.25 (327.27)	1078.12 (426.53)	2.24 (0.36)	1225 (512.34)	746.79 (329.23)	750.00 (427.56)	976.41 (341.12)	2797.06 (1086.88)	4857.18 (1955.01)	2.73 (3.42)	2.8 (2.25)	2.90 (2.35)
Bark														
Infant	843.47 (140.28)	14401.02 (3656.81)	1453.12 (235.93)	2484.37 (815.49)	2.83 (0.47)	3281.25 (1542.37)	1687.5 (153.09)	1453.12 (235.93)	2071.51 (424.88)	4838.68 (1121.20)	7623.77 (1551.54)	0.39 (0.30)	1.25 (0.5)	3.87 (1.65)

Appendices

Juvenile	444.1 (N/A)	16812 (N/A)	1500 (N/A)	2250 (N/A)	2.88 (N/A)	4312.5 (N/A)	1875.00 (N/A)	1500.00 (N/A)	2118.47 (N/A)	5185.43 (N/A)	8806.28 (N/A)	1.39 (1.06)	5.00 (N/A)	2.87 (N/A)
Squeak														
Infant	369.1 (233.44)	8995 (6336.44)	843.75 (446.33)	1453.12 (815.49)	2.85 (1.17)	3000.00 (1785.35)	1125.00 (405.04)	1031.25 (324.75)	1550.24 (1335.49)	3983.13 (2971.55)	5827.55 (4674.50)	0.67 (1.18)	2.2 (2.16)	12.49 (9.49)
Juvenile	50 (N/A)	1063.5 (N/A)	187.5 (N/A)	562.5 (N/A)	1.57 (N/A)	562.5 (N/A)	187.5 (N/A)	187.5 (N/A)	181.04 (N/A)	560.41 (N/A)	671.57 (N/A)	0.18 (N/A)	1.00 (N/A)	5.46 (N/A)

Note: Hz = Hertz, s = Second, Freq = Frequency

I) Means and standard deviation of call acoustics for each cluster of the best fitting infant and juvenile models and their counterpart models.

Descriptive statistics (Mean and SD) of call cluster acoustic characteristics								
	Model parameters: K=2, $\mu=1.1$				Model parameters: K=2, $\mu=1.5$			
Acoustic parameter	Cluster 1 (Infants)	Cluster 1 (Juveniles)	Cluster 2 (Infants)	Cluster 2 (Juveniles)	Cluster 1 (Infants)	Cluster 1 (Juveniles)	Cluster 2 (Infants)	Cluster 2 (Juveniles)
Lowest frequency (Hz)	612.13 (1040.88)	378.44 (660.00)	158.56 (218.55)	185.25 (210.47)	618.22 (1055.29)	388.87 (677.85)	174.33 (250.96)	187.91 (208.63)
Delta frequency (Hz)	12430.35 (3549.68)	12702.19 (3834.73)	2713.66 (3329.52)	5773.15 (4355.69)	12630.35 (3399.65)	12768.41 (3809.19)	2949.11 (3466.14)	6308.46 (4653.63)
Peak frequency (Hz)	1882.34 (1699.68)	1535.53 (1235.13)	524.52 (336.22)	1232.13 (2678.53)	1916.17 (1719.76)	1552.71 (1269.42)	548.75 (356.42)	1235.17 (2554.22)
Q1 frequency (Hz)	1493.58 (1329.25)	1175.70 (787.90)	393.39 (270.19)	626.22 (371.32)	1514.35 (1347.00)	1185.22 (807.00)	422.81 (305.38)	662.71 (387.84)
Q3 frequency (Hz)	3346.57 (1975.84)	3049.50 (1232.80)	724.63 (384.04)	1297 (690.99)	3396.80 (1988.01)	3123.49 (1222.78)	793.67 (511.22)	1351.47 (710.91)
Center frequency (Hz)	2294.73 (1847.18)	1860.96 (925.81)	546.23 (296.73)	875.58 (454.44)	2330.22 (1869.04)	1887.34 (941.03)	589.34 (362.15)	927.47 (485.93)
Mean entropy	3.79 (0.69)	3.85 (0.70)	1.92 (0.60)	2.54 (0.75)	3.82 (0.68)	3.87 (0.67)	1.99 (0.65)	2.63 (0.83)
90% bandwidth (Hz)	4962.65 (2015.14)	5216.02 (2017.83)	964.52 (829.87)	1747.29 (1195.25)	5033.65 (2001.24)	5348.99 (1950.32)	1078.04 (986.09)	1872.67 (1337.09)
F1 (Hz)	2293.96 (673.71)	2333.01 (726.50)	441.35 (285.79)	919.69 (543.81)	2332.29 (653.44)	2385.17 (711.29)	485.96 (339.94)	973.65 (558.70)
F2 (Hz)	4898.14 (1147.68)	4935.11 (1283.84)	929.15 (851.27)	2031.07 (1273.70)	4986.97 (1066.75)	5033.27 (1235.62)	1014.82 (906.64)	2154.84 (1311.89)

Appendices

<i>F3 (Hz)</i>	7609.74 (1682.52)	7619.47 (1964.54)	1474.58 (1418.45)	3192.26 (2080.55)	7738.38 (1561.16)	7752.00 (1903.23)	1620.94 (1526.20)	3405.41 (2158.93)
<i>Duration (s)</i>	2.21 (3.65)	3.01 (4.34)	2.12 (3.43)	8.98 (20.51)	2.20 (3.58)	2.88 (4.40)	2.14 (3.55)	8.62 (19.59)
<i>Element number</i>	5.42 (8.29)	7.53 (10.05)	5.20 (6.52)	16.18 (32.90)	5.24 (7.39)	7.33 (10.21)	5.48 (8.03)	15.68 (31.43)
<i>Element rate (s⁻¹)</i>	8.70 (17.55)	6.49 (9.14)	7.31 (8.72)	7.34 (17.52)	8.45 (17.26)	6.65 (9.32)	7.74 (10.21)	7.03 (16.79)

Note: Hz = Hertz, s = seconds

Appendix 2 – Supplementary results (Chapter 3)

A) Direct comparison of expression patterns between infants and juveniles

Grunts

Mann-Whitney U tests were used to test whether there were significant differences between infants and juveniles in the percentage of their grunts that were associated with positive, neutral, or negative affective states. No significant differences between infants and juveniles were found in the percentage grunts that were associated with positive ($U=0.335$, $N=21$, $p=.747$), neutral ($U=-1.469$, $N=21$, $p=.146$), or negative ($U=-1.374$, $N=21$, $p=.186$) affective states.

Laughs

Mann-Whitney U tests were used to test whether there were significant differences between infants and juveniles in the percentage of their laughs that were associated with positive, neutral, or negative affective states. No significant differences between infants and juveniles were found in the percentage of laughs that were associated with positive ($U=-0.108$, $N=17$, $p=.963$), neutral ($U=-1.015$, $N=17$, $p=.423$), or negative ($U=-1.374$, $N=17$, $p=.481$) affective states.

Screams

Mann-Whitney U tests were used to test whether there were significant differences between infants and juveniles in the percentage of their screams that were associated with neutral or negative affective states. None of the observed screams were related to

positive affect and therefore a comparison between infants and juveniles in the percentage of positive screams were not possible. No significant differences between infants and juveniles were found in the percentage of laughs that were associated with neutral ($U=-1.516$, $N=15$, $p=.328$) or negative ($U=-1.516$, $N=15$, $p=.328$) affective states.

Whimpers

Mann-Whitney U tests were used to test whether there were significant differences between infants and juveniles in the percentage of their whimpers that were associated with positive, neutral, or negative affective states. No significant differences between infants and juveniles were found in the percentage of whimpers that were associated with positive ($U=-0.992$, $N=21$, $p=.554$), neutral ($U=-0.647$, $N=21$, $p=.554$), or negative ($U=-0.719$, $N=21$, $p=.508$) affective states.

Hoo calls

There was insufficient data on infants ($N=4$) to perform direct infant vs juvenile comparisons of hoo calls.

B) Direct comparisons of call functionality between infants and juveniles, and functional flexibility analyses for laughs, screams, and whimpers

Grunts

Direct comparisons of infants vs juveniles using Mann-Whitney U tests showed that there was no significant difference in the percentage of negative grunts that elicited behavioural change in a social partner ($U=1.760$, $N=20$, $p=.092$). Similarly, there was no

significant difference in the percentage of negative grunts that elicited behavioural continuation in a social partner between infants and juveniles ($U=-1.678$, $N=20$, $p=.092$). For positive grunts, there was no significant difference between infants and juveniles in the percentage of grunts that elicited behavioural change ($U=.173$, $N=20$, $p=.865$) or continuation ($U=.173$, $N=20$, $p=.865$) in a social partner. Similarly, there was no significant difference between infants and juveniles for neutral grunts in the percentage that elicited behavioural change ($U=-1.734$, $N=20$, $p=.083$) or continuation ($U=-1.734$, $N=20$, $p=.083$) in a social partner.

Laughs

A Friedman test showed an overall significant differences in the affective valence of laughs that elicited behavioural change in a social partner, $\chi^2(3)=18.750$, $N=17$, $p<.001$. Wilcoxon post-hoc comparisons showed that when a laugh elicited behavioural change in a social partner, it was significantly more likely to be positive than either neutral ($Z=-3.354$, $N=17$, $p=.002$) or negative ($Z=-3.354$, $N=20$, $p=.001$), but there was no significant difference in the percentage of neutral vs negative laughs that elicited behavioural change in a social partner ($Z=.000$, $N=17$, $p=.333$).

When infants were examined alone, a Friedman test showed a significant difference in the affective valence of laughs that elicited behavioural change in a social partner, $\chi^2(3)=10.000$, $N=8$, $p=.006$. Wilcoxon post-hoc tests showed no significant difference in the percentage of positive vs neutral ($Z=-2.372$, $N=8$, $p=.053$) or neutral vs negative laughs that were associated with behavioural changes in social partners ($Z=-.000$, $N=8$, $p=1.000$).

However, positive vs negative laughs ($Z=-2.372$, $N=8$, $p=.026$) were more likely to be associated with behavioural changes in social partners.

When juveniles were examined alone, a Friedman test showed that there was a significant difference in the percentage of positive, neutral, and negative laughs that elicited behavioural change in a social partner, $\chi^2(3)=8.824$, $N=9$, $p=.002$. Wilcoxon post-hoc tests showed that positive laughs were significantly more likely to elicit behavioural changes in a social partner than either neutral ($Z=-2.598$, $N=9$, $p=.015$) or negative ($Z=-2.598$, $N=9$, $p=.028$) laughs. However, there was no significant difference in the percentage of neutral vs negative laughs that elicited behavioural change in a social partner ($Z=.000$, $N=9$, $p=.333$).

Mann-Whitney U tests were used to test whether a greater percentage of positive, neutral, or negative laughs elicited behavioural change in social partners in infants compared to juveniles. These tests showed that there was no significant difference in the proportion of positive ($U=-1.217$, $N=17$, $p=.476$), neutral ($U=-0.816$, $N=17$, $p=.762$), or negative ($U=-0.816$, $N=17$, $p=.762$) laughs that elicited behavioural change in a social partner between infants and juveniles.

Overall, a Friedman test showed that there was a significant difference in the percentage of positive, neutral, and negative laughs that were associated with behavioural continuation in a social partner, $\chi^2(3)=14.387$, $N=17$, $p<.001$. Wilcoxon post-hoc comparisons showed that positive laughs were significantly more likely to be associated with behavioural continuation in a social partner than either neutral ($Z=-2.571$, $N=17$,

$p=.015$) or negative ($Z=-3.130$, $N=17$, $p=.005$) laughs. However, there was no significant difference in the percentage of neutral vs negative laughs that elicited behavioural continuation in a social partner ($Z=-.559$, $N=17$, $p=.333$). Figure S1 below depicts the mean and standard error of the percentage of laugh vocalisations that were associated with social partners changing or continuing their behaviour as a function of laugh valence.

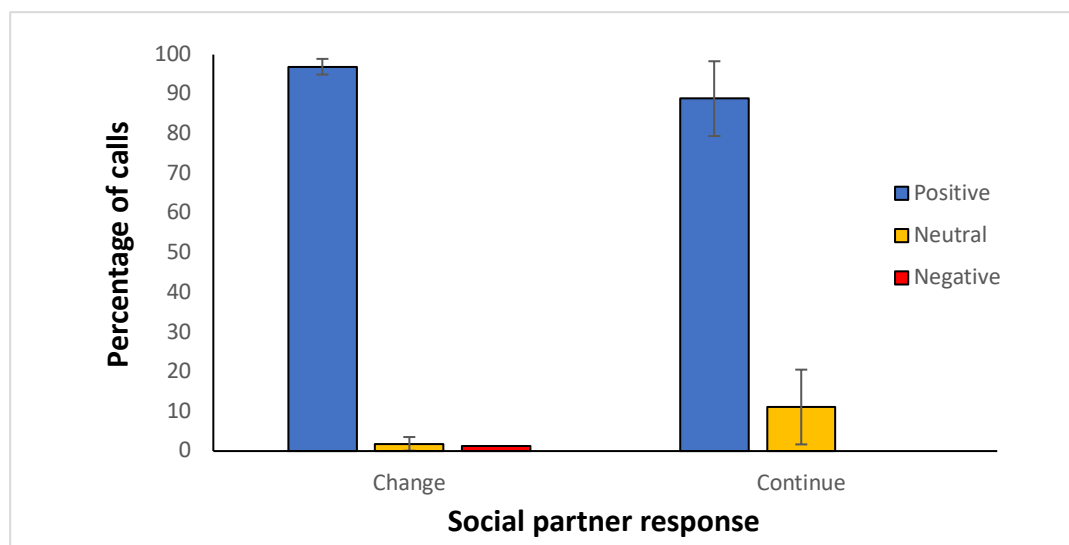


Figure S1. Bar graph depicting the mean and standard error of the percentage of positive, neutral and negative laughs where social partners changed or continued their behaviour.

Only 4 infants ever produced laughs associated with behavioural continuation in a social partner which was insufficient for statistical analysis. However, laughs were associated with behavioural continuation in a social partner much more often among juveniles and it was possible to test whether there was a significant difference in the percentage of positive, neutral, and negative laughs that were associated with behavioural continuation in a social partner. A Friedman test showed that in juveniles there was a significant difference in the percentage of positive, neutral, and negative laughs that were

associated with behavioural continuation in a social partner, $\chi^2(3)=11.474$, $N=9$, $p=.001$. Wilcoxon post-hoc tests showed that positive laughs were significantly more likely to be associated with behavioural continuation in a social partner than neutral ($Z=-2.454$, $N=9$, $p=.021$) or negative ($Z=-2.742$, $N=9$, $p=.018$) laughs. However, there was no significant difference in the percentage of neutral vs negative laughs associated with behavioural continuation in a social partner ($Z=-.289$, $N=9$, $p=.333$). Given the small number of infants that used laughs to elicit behavioural continuation in a social partner, comparisons between infants and juveniles in the percentage of positive, neutral, and negative laughs that were associated with behavioural continuation in a social partner were not possible.

Comparisons within-subjects of the percentage of positive laughs that elicited behavioural change vs continuation showed that there was no significant difference overall ($Z=-.889$, $N=17$, $p=.393$) or for juveniles alone ($Z=-.420$, $N=9$, $p=.723$). An infant only analysis was not possible due to insufficient data. Similarly, an analysis for neutral and negative laughs was not possible due to the small number of such calls that occurred.

Screams

Overall, a Friedman test showed that there was a significant difference in the percentage of positive, neutral, and negative screams that were associated with behavioural change in a social partner, $\chi^2(3)=18.500$, $N=15$, $p<.001$. Wilcoxon post-hoc comparisons showed that negative screams were significantly more likely to elicit behavioural change in a social partner than either neutral ($Z=3.062$, $N=15$, $p=.003$) or positive ($Z=3.368$, $N=15$, $p=.002$) screams. However, there was no significant difference in the percentage of positive vs neutral screams that elicited behavioural change in a social partner ($Z=.306$,

N=15, $p=.333$). Figure S2 below depicts the mean and standard error of the percentage of scream vocalisations that were associated with social partners changing their behaviour as a function of scream valence.

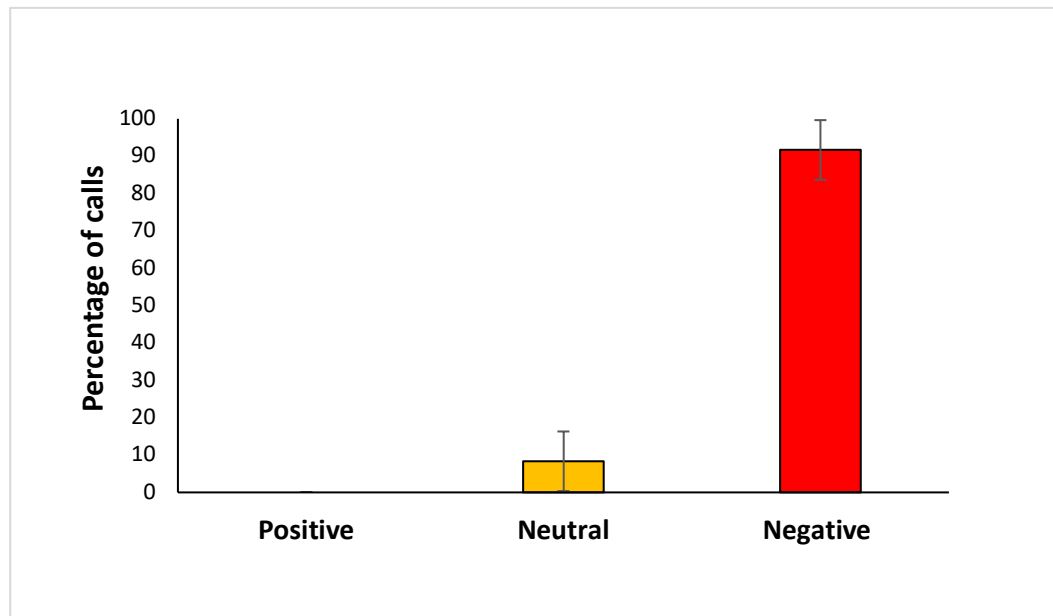


Figure S2. Bar graph depicting the mean and standard error of the percentage of positive, neutral and negative laughs where social partners changed their behaviour.

Only 4 infants were observed to use screams to elicit behavioural change in a social partner. Consequently, there was insufficient data for an infant-only analysis. In contrast, most juveniles used screams to elicit behavioural change in a social partner.

Consequently, it was possible to use a Friedman test to compare the percentage of positive, neutral, and negative screams that elicited behavioural change in a social partner for juveniles. This test showed that there was a significant difference in the percentage of positive, neutral, and negative screams that elicited behavioural change in a social partner among juveniles, $\chi^2(3)=12.667$, $N=9$, $p=.002$. Wilcoxon post-hoc

comparisons showed that negative screams were significantly more likely to elicit behavioural change in a social partner than positive screams ($Z=2.625$, $N=9$, $p=.014$) and neutral screams ($Z=2.250$, $N=9$, $p=.025$). However, there was no significant differences in the percentage of neutral vs positive screams that elicited behavioural change in a social partner ($Z=.188$, $N=9$, $p=.333$)

Given the small amount of observations of infants using screams to elicit behavioural change in a social partner, comparisons between infants and juveniles in the percentage of positive, neutral, and negative screams that elicited behavioural change in a social partner were not possible. Moreover, screams were observed to elicit a social partner to continue a behaviour on only a single occasion. Consequently, no statistical analysis relating to behavioural continuation in a social partner was possible. This may suggest that screams are more functionally stereotyped than other call types that elicited a sufficient number of different types of behavioural responses for statistical analyses.

Whimpers

Overall, a Friedman test showed that there was a significant difference in the percentage of positive, neutral, and negative whimpers that were associated with behavioural change in a social partner, $\chi^2(3)=12.286$, $N=21$, $p=.001$. Wilcoxon post-hoc comparisons showed that significantly more negative whimpers were associated with behavioural change in a social partner than either neutral ($Z=2.740$, $N=21$, $p=.018$) or positive ($Z=2.646$, $N=21$, $p=.012$) whimpers. However, there was no significant difference in the percentage of positive vs neutral grunts that were associated with behavioural change in a social partner ($Z=-.094$, $N=21$, $p=.333$). Figure S3 below depicts the mean and standard error of

the percentage of whimper vocalisations that were associated with social partners changing their behaviour as a function of whimper valence.

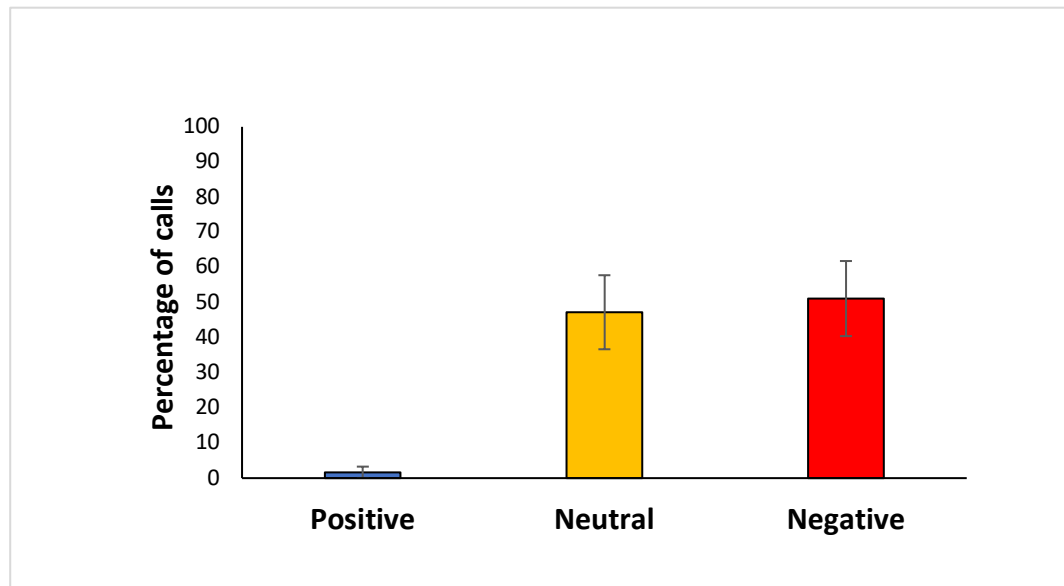


Figure S3. Bar graph depicting the mean and standard error of the percentage of positive, neutral and negative whimpers where social partners changed their behaviour.

When infants were examined alone, a Friedman test showed that there was a significant difference in the percentage of positive, neutral, and negative whimpers that elicited behavioural change in a social partner, $\chi^2(3)=9.548$, $N=12$, $p=.004$. Wilcoxon post-hoc comparisons showed that the percentage of negative whimpers that elicited behavioural change in a social partner was significantly higher than for positive whimpers ($Z=2.593$, $N=12$, $p=.029$) and neutral whimpers ($Z=2.357$, $N=12$, $p=.027$). However, there was no significant difference in the percentage of neutral vs positive whimpers that elicited behavioural change in a social partner ($Z=.111$, $N=12$, $p=.333$).

In contrast to infants, a Friedman test found that there was no significant difference in the percentage of positive, neutral, and negative whimpers that elicited behavioural change in a social partner for juveniles, $\chi^2(3)=2.800$, $N=9$, $p=.128$. Mann-Whitney U tests showed that there was no significant difference between infants and juveniles in the percentage of positive ($U=-0.745$, $N=9$, $p=.787$), neutral ($U=-0.343$, $N=9$, $p=.787$), or negative ($U=0.342$, $N=9$, $p=.266$) whimpers that elicited behavioural change in a social partner. Since whimpers mostly elicited behavioural change, comparisons of the percentage of whimpers that elicited change vs continuation as a function of valence were not possible.

Hoo calls

Only 3 hoo calls occurred during social interactions and therefore functional flexibility analyses were not possible.

Appendix 3 – Supplementary methods & results (Chapter 4)

Supplementary methods

A) Main model diagnostics

To check for co-linearity between predictors, the variance inflation factor was examined, which did not exceed 5 for any predictors in any model, suggesting independence of predictor variables (Fox & Monette, 1992). Cooks distance values were examined to identify observations that appeared to have a large influence on the model. Values where cooks distance $>3 \times \text{mean cooks distance}$ were inspected (Cook, 1979). These values were only found in the models where mutual gaze and gaze directedness were the dependent variables where 1 and 2 such values were found respectively. Upon inspection it was

decided there was nothing unusual about these values and they were therefore kept in the models. We tested for over dispersion using the DHARMA package V0.3.3.0 in R (Hartig, 2017) which showed none of the models showed signs of over dispersion. Finally, to examine the assumption of linearity between continuous predictors and the logit of the dependent variable, age (which was the only continuous predictor) was plotted against the logit of the dependent variable for all models wherein age was a predictor, and a loess curve was fitted to the data. Examination of loess curves fitted to the data showed that the relationship between age and the logit of the dependent variable was linear in all models. As such, we conclude that all the main models are a good fit to the data.

Gaze directed main model diagnostics

Table S1. Variance inflation factor values for gaze directed model. Relevant values according to Fox & Monette (1972) are in bold.			
	GVIF	df	GVIF ^{(1/(2*df))}
Age	1.21	1	1.10
Call Type	63.97	3	1.99
Age*Call Type	68.90	3	2.02

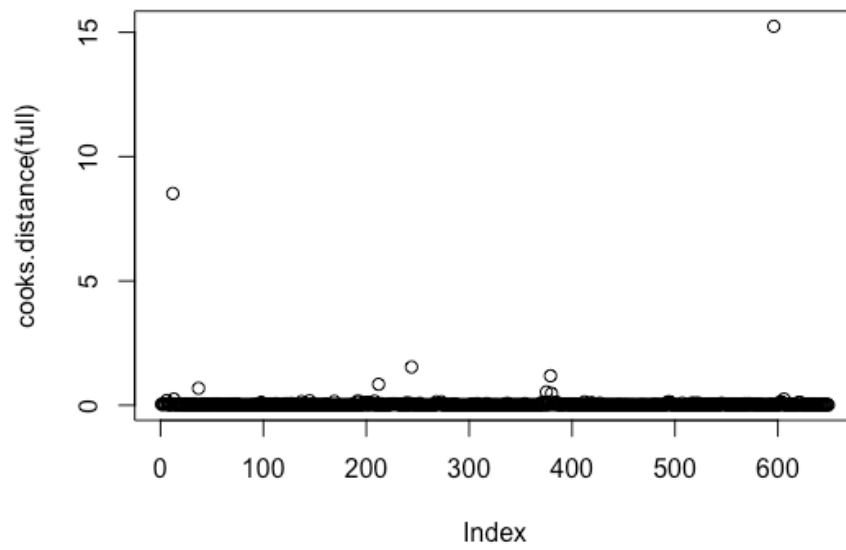


Figure S1. Cooks distance values for the gaze directed model

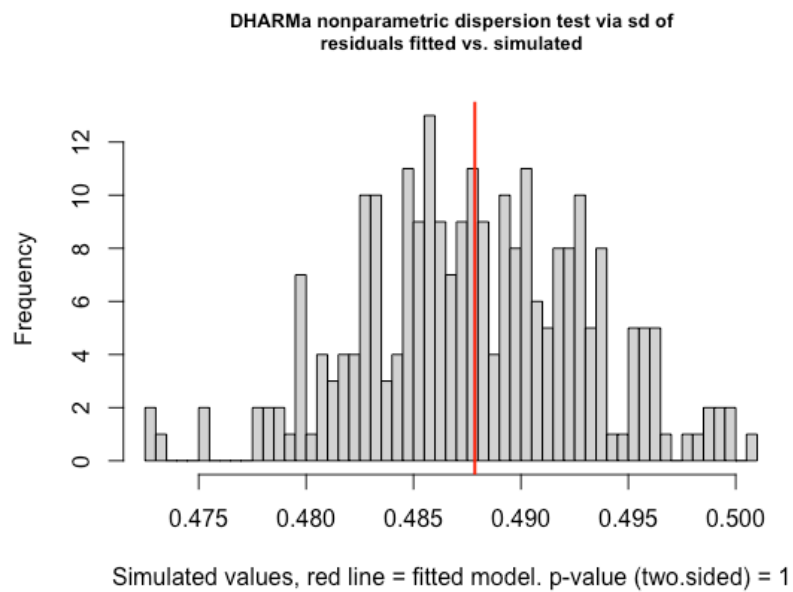


Figure S2. Dispersion test for gaze directed model showing no significant over/under dispersion.

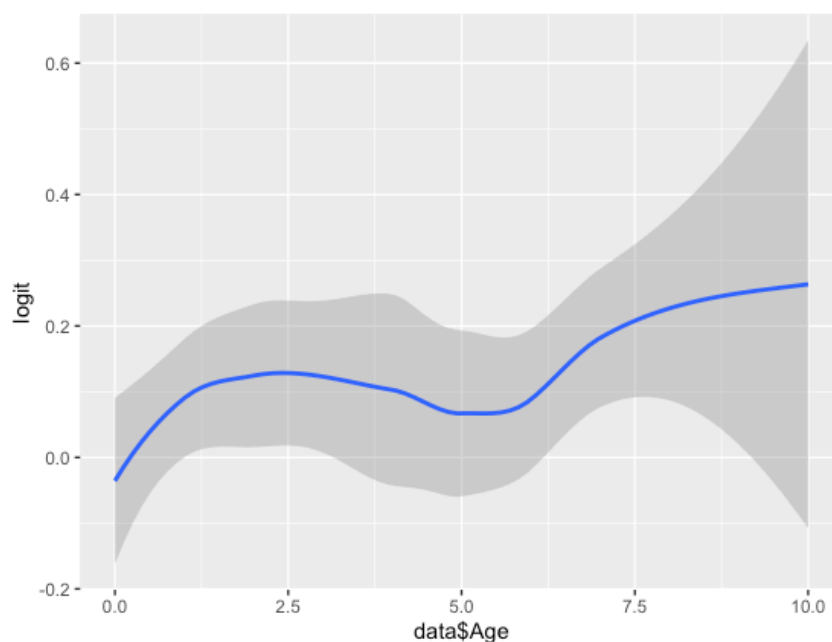


Figure S3. Age plotted against the logit of directedness in the gaze directed model with a fitted loess curve showing a linear relationship.

Face directed model diagnostics

Table S2. Variance inflation factor values for face directed model. Relevant values according to Fox and Monette (1972) are in bold.			
	GVIF	df	$GVIF^{(1/(2*df))}$
Age	1.32	1	1.14
Call Type	47.68	3	1.90
Age*Call Type	51.13	3	1.92

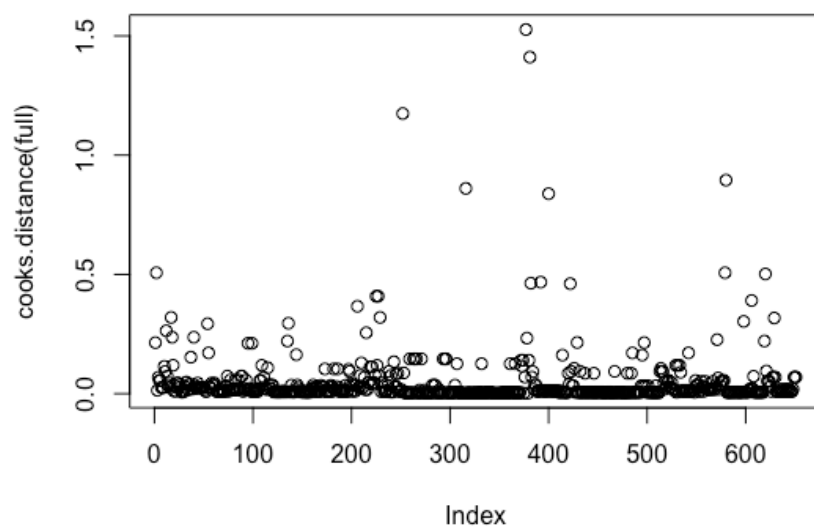


Figure S4. Cooks distance values for the face directed model

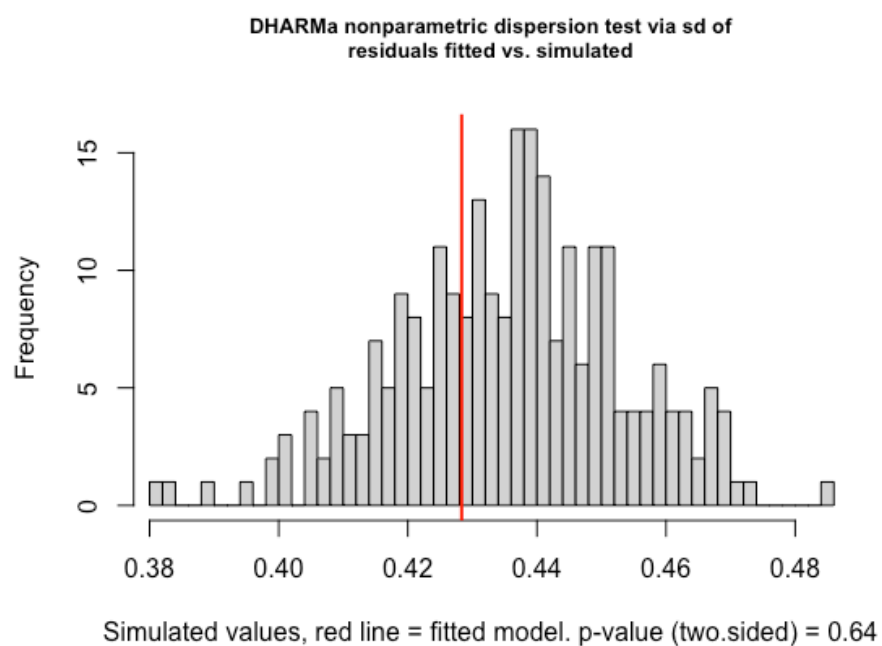


Figure S5. Dispersion test for face directed model showing no significant over/under dispersion.

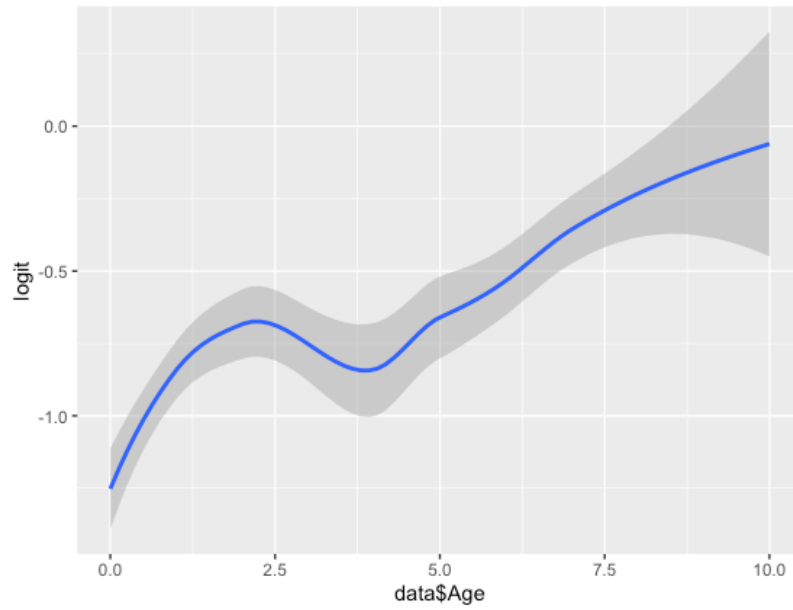


Figure S6. Age plotted against the logit of directedness in the face directed model with a fitted loess curve showing a linear relationship.

Mutual gaze model diagnostics

Table S3. Variance inflation factor values for mutual gaze model. Relevant values according to Fox and Monette (1972) are in bold.			
	GVIF	df	GVIF ^{(1/(2*df))}
Age	1.35	1	1.16
Call Type	55.88	3	1.95
Age*Call Type	59.55	3	1.97

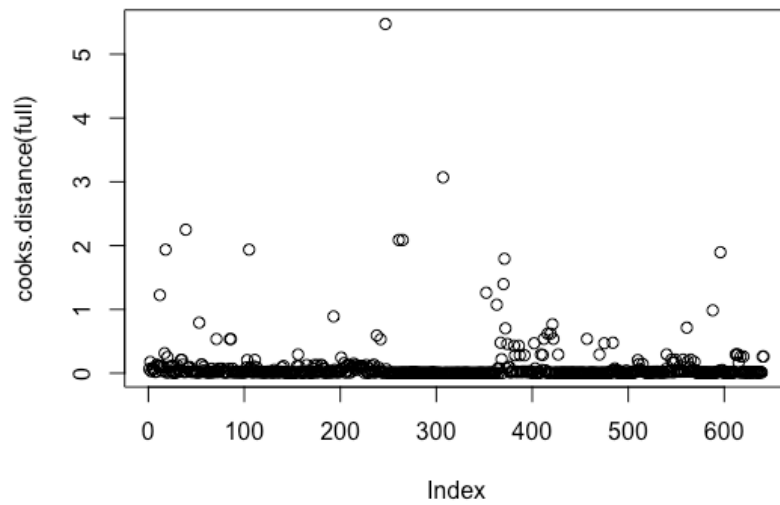


Figure S7. Cooks distance values for mutual gaze model

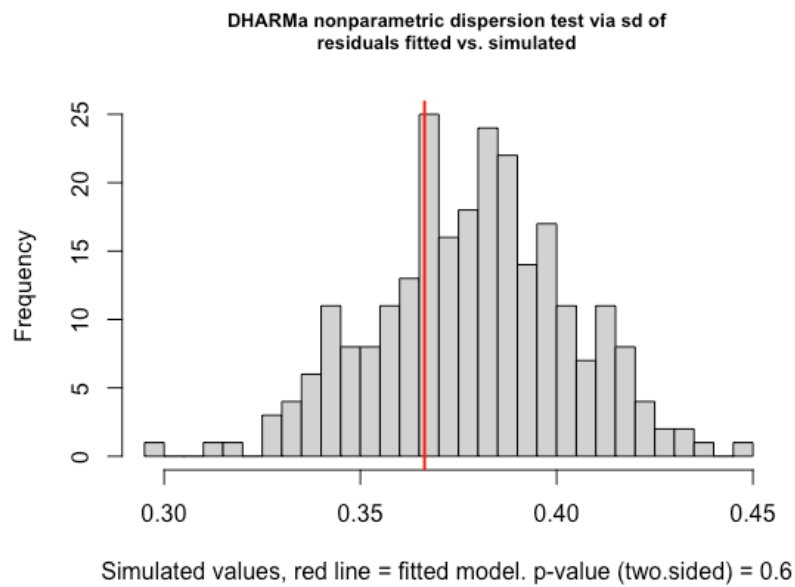


Figure S8. Dispersion test for mutual gaze model showing no significant over/under dispersion.

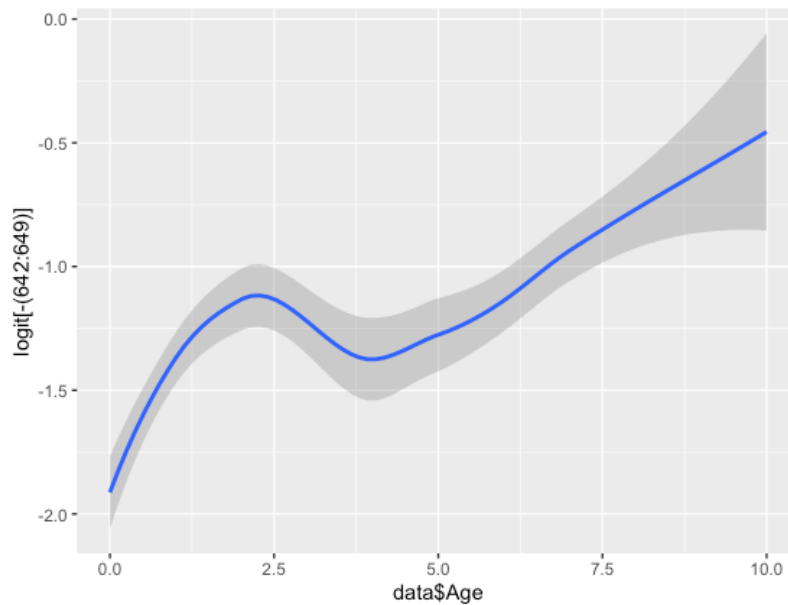


Figure S9. Age plotted against the logit of directedness in the mutual gaze model with a fitted loess curve showing a linear relationship.

B) Function model diagnostics

There are no VIF values due to there being a single predictor and therefore no possibility of multi-collinearity. There is no test of the linearity assumption because the predictor is a binary variable.

Gaze directedness as a predictor of behavioural response in a social partner

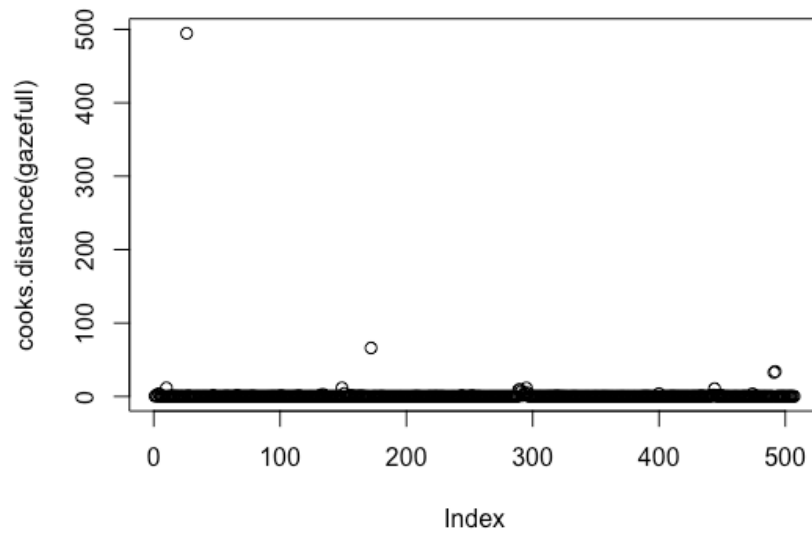


Figure S10. Cooks distance values for gaze as a predictor of behavioural response in a social partner.

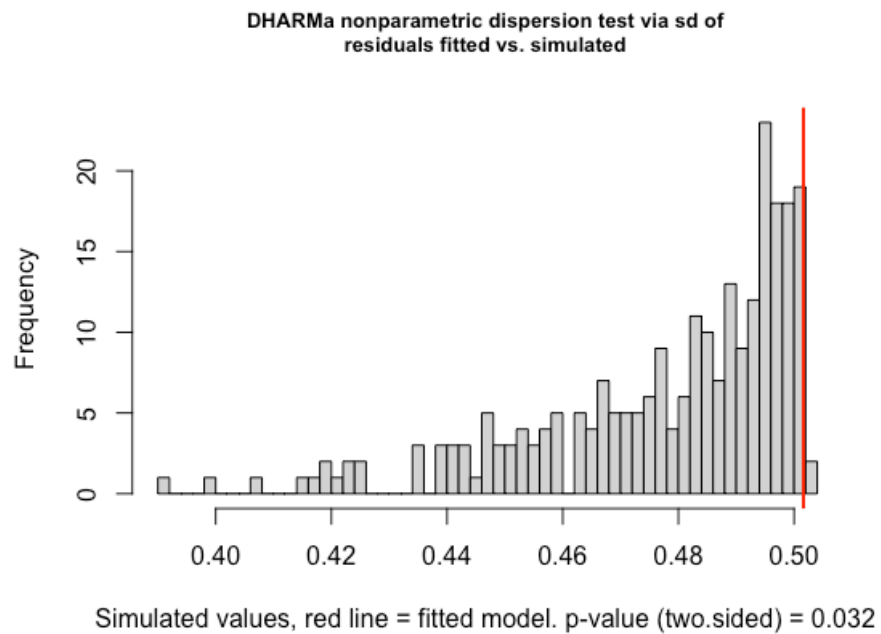


Figure S11. Dispersion test for mutual gaze model showing significant over dispersion.

Face directedness as a predictor of behavioural response from a social partner

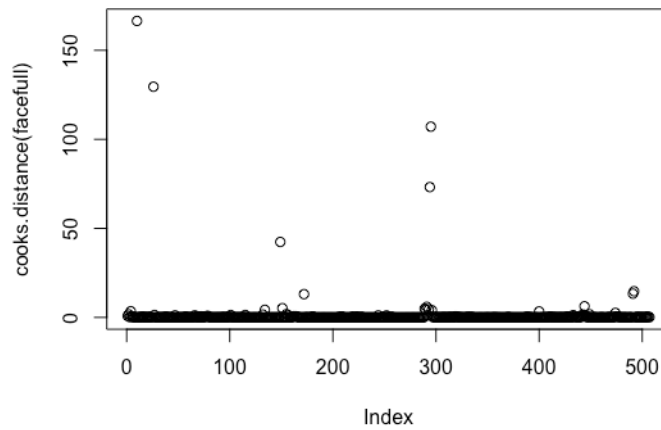


Figure S12. Cooks distance values for face directedness as a predictor of behavioural response in a social partner.

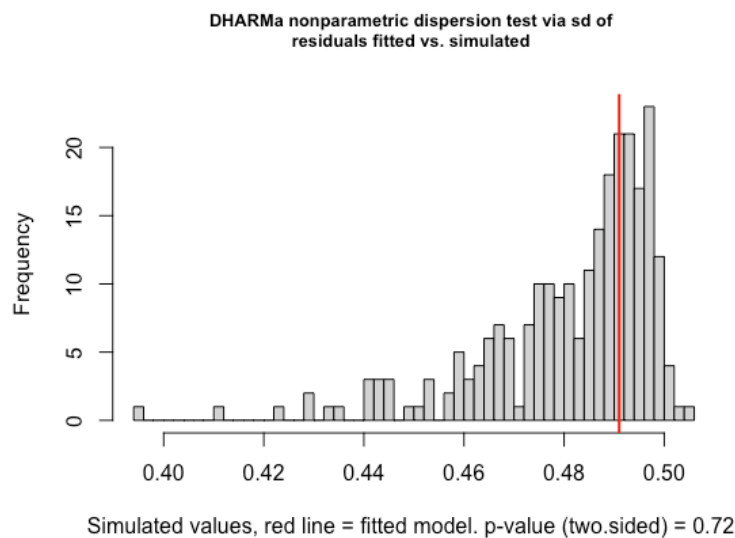


Figure S13. Dispersion test for face directedness as a predictor of behavioural response in a social partner showing no significant over/under dispersion.

Mutual gaze as a predictor of behavioural response from a social partner

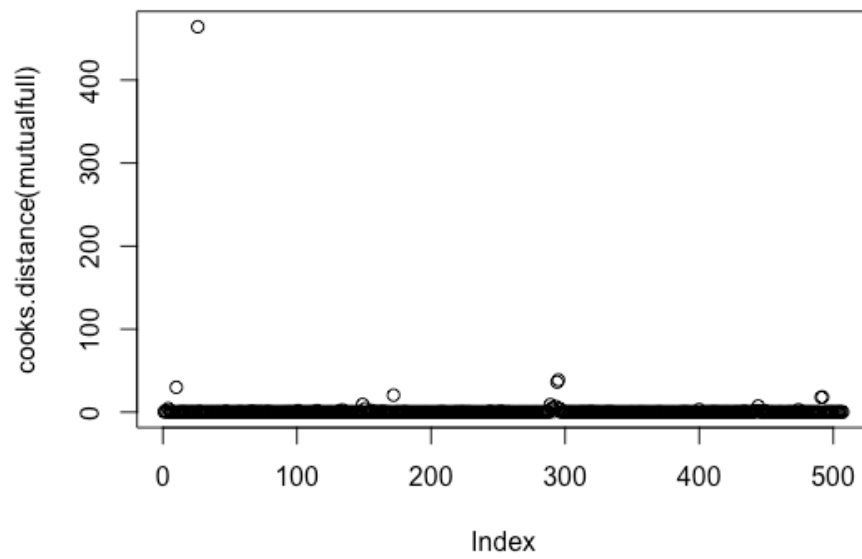


Figure S14. Cooks distance values for mutual gaze as a predictor of behavioural response from a social partner.

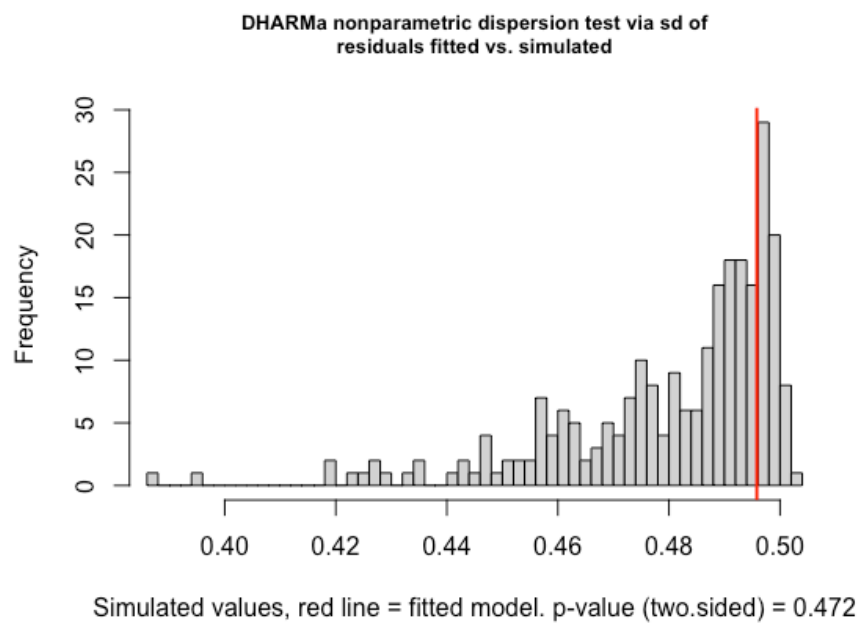


Figure S15. Dispersion test for mutual gaze as a predictor of behavioural response in a social partner showing no significant over/under dispersion.

Supplementary results

A) Gaze directed final model

Table S1. Main effects in the reduced model with fixed effects of age and call type, and post-hoc comparisons with Tukey corrections

Predictors	Estimate	Standard error	z-value	p-value
Age	0.004	0.058	-0.335	0.939
Call type: laugh	0.474	0.225	2.103	0.035
Call type: scream	1.50	0.450	3.343	0.0008
Call type: whimper	0.820	0.244	3.364	0.0007
Grunt - Laugh	-0.474	0.225	-2.103	0.152
Grunt - Scream	-1.508	0.451	-3.343	0.004
Grunt - Whimper	-0.821	0.244	-3.363	0.004
Laugh - Scream	-1.034	0.473	-2.184	0.127
Laugh – Whimper	-0.347	0.293	-1.183	0.637
Scream - Whimper	-0.687	0.476	1.445	0.472

Note: Grunts are taken as the reference category for the fixed effect of call type. Values highlighted in bold are significant at the $p < 0.05$ level

B) Face directed final model**Table S2.** Main effects and interactions for the full model predicting face directedness with the fixed effects of age and call type.

Fixed effects	Estimate	Standard error	z-value	p-value
Age	0.148	0.073	2.020	0.043
Call type: laugh	-0.576	0.586	-0.983	0.325
Call type: scream	1.441	0.803	1.794	0.072
Call type: whimper	1.537	0.414	3.712	0.0002
Grunt - Laugh	0.015	0.313	0.049	1.000
Grunt - Scream	-1.166	0.453	-2.577	0.049
Grunt - Whimper	-0.797	0.274	-2.915	0.018
Laugh - Scream	-1.181	0.511	-2.310	0.095
Laugh - Whimper	-0.813	0.363	-2.238	0.113
Scream - Whimper	0.368	0.471	0.782	0.862
Age * laugh	0.164	0.105	1.560	0.118
Age * scream	-0.080	0.147	-0.548	0.538
Age * whimper	-0.216	0.102	-2.108	0.035
Age * Grunt – Age * Laugh	-0.162	0.106	-1.560	0.401
Age * Grunt – Age * Scream	0.080	0.147	0.548	0.947
Age * Grunt – Age * Whimper	0.216	0.103	2.108	0.150
Age * Laugh – Age * Scream	0.245	0.168	1.458	0.462
Age * Laugh – Age * Whimper	0.381	0.133	2.861	0.022

Age * Scream – Age * Whimper 0.136 0.160 0.848 0.831

Note: Grunts are taken as the reference category for the fixed effect of call type. Values highlighted in bold are significant at the $p < 0.05$ level

C) Mutual gaze final model

Table S3. Main effects and interactions for the full model predicting mutual gaze with the fixed effects of age and call type.

Fixed effects	Estimate	Standard error	z-value	p-value
Age	0.130	0.088	1.484	0.137
Call type: laugh	-0.815	0.717	-1.136	0.255
Call type: scream	3.105	0.930	3.338	0.0008
Call type: whimper	1.433	0.499	2.871	0.004
Grunt – Laugh	0.069	0.388	0.178	0.998
Grunt – Scream	-1.710	0.513	-3.335	0.004
Grunt – Whimper	-0.791	0.326	-2.427	0.072
Laugh – Scream	-1.779	0.593	-3.002	0.014
Laugh – Whimper	-0.860	0.440	-1.955	0.205
Scream - Whimper	0.919	0.533	-1.725	0.310
Age * laugh	0.219	0.122	1.790	0.073
Age * scream	-0.409	0.176	-2.324	0.020
Age * whimper	-0.188	0.125	-1.504	0.132
Age * Grunt – Age * Laugh	-0.219	0.122	-1.790	0.277

Age * Grunt – Age * Scream	0.410	0.176	2.324	0.092
Age * Grunt – Age * Whimper	0.189	0.125	1.504	0.434
Age * Laugh – Age * Scream	0.629	0.200	3.145	0.009
Age * Laugh – Age * Whimper	0.408	0.159	2.558	0.052
Age * Scream – Age * Whimper	-0.221	0.191	-1.154	0.655

Note: Grunts are taken as the reference category for the fixed effect of call type. Values highlighted in bold are significant at the $p < 0.05$ level

D) KR20 final model

Table S4. Main effects in the final model using age and call type as predictors of KR20 values.

Fixed effects	Estimate	Standard error	t-value	p-value
Age	0.044	0.076	6.456	0.001
Call type: laugh	-0.075	0.089	-0.835	0.403
Call type: whimper	-0.003	0.115	-12.484	0.974

Note: Grunts are taken as the reference category for the fixed effect of call types and screams were not included because there were no observed screams among the subjects for whom KR20 values could be calculated. Values highlighted in bold are significant at the $p < .05$ level.

Appendix 4 – AWERB ethical approval

14 May 2018

Dear Mr Taylor and Dr Davila-Ross,

RE: Ethics submission – **The Development of Vocal Functional Flexibility in Semi-Wild Chimpanzees (*Pan troglodytes*)**.

Approval of project by the Animal Welfare and Ethical Review Body (AWERB)

I am very happy to confirm that we were able to fast track your application and that the AWERB gave its approval for your proposal concerning work within the above project.

The AWERB uses UK Home Office guidelines on the Animals (Scientific Procedures) Act 1986 when assessing proposals and adheres to the regulations of the European Directive 2010/63/EU. Your project has been assessed as not causing pain, suffering or lasting harm to the animal and so falls outside the A(SP)A. We are confident that the proposal demonstrates appropriate consideration of the Three Rs and animal welfare. Please use this letter as confirmation of ethical approval from AWERB, University of Portsmouth. Please use the number 818B as confirmation of the successful review.

Yours sincerely,

MJ Guille PhD FSB

Professor of Developmental Genetics and Chair, AWERB

A handwritten signature in black ink, appearing to be 'MJ Guille', written in a cursive style.

Appendix 5 – Chimfunshi Wildlife Orphanage research approval

March, 2018

Review of Chimfunshi Research Advisory Board Proposal

Project title: “Free expressivity in hominin vocalizations”

Applicant: Derry Taylor, Marina Davila-Ross

CRAB collaborator: Edwin van Leeuwen

Dear team,

The Chimfunshi Research Advisory Board has completed the review of your proposal. We review the proposals for practicality/feasibility, chimpanzee welfare, human and chimpanzee safety, and scientific rigor. You can find minor comments relating to each of these topics below (when applicable). We are pleased to tell you that we do not have any major concerns about hosting this project at Chimfunshi, and are glad that you are planning to do this work at Chimfunshi. Please provide us a response to the comments below so that we can have a final approval on file.

We look forward to hearing from you, and to hosting your project!

Sincerely,

The Chimfunshi Research Advisory Board

Practicality/feasibility:

This project is feasible and significant and has ambitious aims (achievable?). As purely observational, minimal disruption of daily operations will be necessary. Minimal keeper input will be only necessary during the animal ID phase, but overall will be self-run.

Chimpanzee welfare:

No concerns.

Human & chimpanzee safety:

For safety, no interaction with chimpanzees should be allowed. The researcher after briefed by one the keepers/staff should stay in the designated safe areas while indoors, and always preferably accompanied by the area keeper. If not possible, keepers should always be informed when researchers hope to film indoors and when out.

Scientific rigor:

No concerns.

March, 2018

Other comments:

This review pertains only to the research proposed for chimpanzees. Research regarding human subjects is not considered by the Chimfunshi Research Advisory Board (please ask Edwin how to best go about this).

Please, be aware that during the selected period, there is a possibility of ongoing new animal introduction process of orphan animals of that age range that might complicate or interfere with this research. This will be discussed timely with the researcher if possibility confirmed.

Also, please be aware of the fact that other researchers may be conducting their studies at the same time as you will. We will, however, keep this to a minimum such that all studies can be properly conducted.

Appendix 6 – UPR16 Form

FORM UPR16**Research Ethics Review Checklist**

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information)

Postgraduate Research Student (PGRS) Information		Student ID:	UP690042
PGRS Name:	Derry Taylor		
Department:	Psychology	First Supervisor:	Marina Davila-Ross
Start Date: (or progression date for Prof Doc students)	01.10.2017		
Study Mode and Route:	Part-time <input type="checkbox"/> Full-time <input checked="" type="checkbox"/>	MPhil <input type="checkbox"/> PhD <input checked="" type="checkbox"/>	MD <input type="checkbox"/> Professional Doctorate <input type="checkbox"/>
Title of Thesis:	The Infra-Structure of Chimpanzee Vocal Development		
Thesis Word Count: (excluding ancillary data)	41,112		
<p>If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study</p> <p>Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).</p>			
UKRIO Finished Research Checklist:			
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)			
a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>		
b) Have all contributions to knowledge been acknowledged?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>		
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>		
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>		
e) Does your research comply with all legal, ethical, and contractual requirements?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>		
Candidate Statement:			
I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)			
Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):		818E	
If you have <i>not</i> submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:			
[]			
Signed (PGRS):	Derry Taylor		Date: 08.12.2020

